National Library of Canada
Canadian Theses Service

Ottawa, Canada
K1A 0N4

Bibliothèque nationale du Canada
Service des thèses canadiennes

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

NL-339 (r. 88/04) c

Canada

National Library of Canada
Canadian Theses Service

Bibliothèque nationale du Canada
Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-315-55652-8

Canada

THE UNIVERSITY OF ALBERTA

THE LINGUISTIC RELATIVITY HYPOTHESIS

by

DONALD H. MOTTERSHEAD

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

DEPARTMENT .....SOCIOLOGY...............

EDMONTON, ALBERTA

FALL, 1989

THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: .....Don Mottershead.............................

NAME OF THESIS: .....The Linguistic Relativity Hypothesis......

....

....

DEGREE: .....Doctor of Philosophy in Sociology................

YEAR THIS DEGREE GRANTED ....Fall, 1989..........................

Permission is hereby granted to THE UNIVERSITY OF ALBERTA LIBRARY to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific purposes only.

The author reserves other publication rites, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

...Don Mottershead...........

PERMANENT ADDRESS:

....3612 Hillview Crescent........

....Edmonton, Alberta, Canada.....

....

DATED ..October 12, 1989...

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled The Linguistic Relativity Hypothesis submitted by Donald H. Mottershead in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Sociology.

Richard Jung

Supervisor

Bernard Linsky

Ragnar Rommetveit

External Examiner

Date 15 August 1989

ABSTRACT

The purpose of this dissertation is to examine the plausibility of the Linguistic Relativity Hypothesis. As formulated by B. L. Whorf, the Linguistic Relativity Hypothesis is the claim that the contents of one's thoughts are strongly conditioned by one's native language. Whorf believed that this conditioning is so powerful that speakers of highly divergent languages cannot understand each other's thoughts.

My thesis is that Whorf is wrong. More specifically, I argue that a basic level of cross-cultural understanding is always possible, and even where there are significant differences in thought, language is never the key factor in explaining the differences.

The argument proceeds by first examining the concept of natural language. In order to express the Linguistic Relativity Hypothesis it is necessary to claim that natural languages have both syntactic and semantic properties. A number of competing approaches to the study of natural language are presented, and the truth-conditional approach is defended as the only one that provides a viable semantic theory. It is argued that the truth-conditional approach is necessarily part of a larger theoretical enterprise, the theory of social action.

Next the concept of thought is considered. It is argued that thought, like natural language, is best understood in the

context of a theory of social action. In this theoretical context, to determine a person's thoughts is to identify his beliefs and desires.

In the context of the theory of social action, the Linguistic Relativity Hypothesis amounts to the claim that we should be able to detect a relation between the beliefs and desires of people and the truth-conditions of the language they speak. However, if we consider the question of how one would determine beliefs, desires and truth-conditions on the basis of the behavioral patterns of a language community, it becomes apparent that it is always possible, and indeed, necessary, to conclude that all actors share a basic stock of fundamental beliefs. Furthermore, it is argued that even where differences of belief exist, these differences can always be attributed to non-linguistic causes.

ACKNOWLEDGEMENTS

I would like to thank Richard Jung, Bernie Linsky, and Leo Mos for their encouragement, support and patience. Anyone who has attempted to complete a thesis while holding down a full-time job knows that at times the task can seem overwhelming. Richard, Bernie, and Leo were instrumental in getting me through those times as well as providing the traditional academic guidance. I am grateful for their help.

Gwynn Nettler was one of the original members of my committee. When he went off campus for an extended period he could no longer formally participate on the committee, but he continued to retain a strong interest in my progress and provided extensive commentary on my drafts. His efforts are appreciated.

Leslie Hayduk replaced Gwynn on the committee. He quickly became acquainted with a work that was by that time nearly complete, and provided a host of fresh insights. I am grateful for his participation and enthusiasm. I am also appreciative of Ragnar Rommetveit for functioning as the external examiner.

Finally, closer to home, my family - Jean, Jeff, and Blair - provided the support and working time that enabled me to complete this task. My mother, Hilda Stewart, also played an instrumental role, for she taught me, at an early age, to value academic achievements. I am thankful to all of them.

TABLE OF CONTENTS

PAGE

VOLUME I

CHAPTER ONE - WHORF'S HYPOTHESIS 1

1.1 WHORF'S HYPOTHESIS 1

1.2 WHORF'S ARGUMENT 7

1.3 THE WESTERN MIND AND THE HOPI MIND 21

1.4 MIND: CONTENT OR CONSCIOUSNESS? 27

1.5 EMPIRICAL RESEARCH 37

1.6 PREVIEW 42

CHAPTER TWO - NATURAL LANGUAGE 46

2.1 INTRODUCTION 46

2.2 BLOOMFIELD'S LINGUISTIC PROGRAM 52

2.3 CONCEPTUAL SEMANTICS 114

2.4 TRUTH-CONDITIONAL SEMANTICS 196

2.5 CONCLUSION 264

VOLUME II

CHAPTER THREE - THOUGHT 271

3.1 THE GEISTESWISSENSCHAFTEN 271

3.2 NON-PROPOSITIONAL THEORIES OF THOUGHT 281

3.3 REDUCIBILITY, SUPERVENIENCE, AND PROJECTIBILITY . 295

3.4 FODOR'S FUNCTIONALISM 317

3.5 AUTONOMOUS SOCIAL SCIENCE 343

CHAPTER FOUR - RELATIVITY 361

4.1 INTRODUCTION 361

4.2 ATTITUDES, MEANING AND PHYSICS 365

4.3 RADICAL INTERPRETATION 399

4.4 DAVIDSON ON CONCEPTUAL RELATIVITY 420

4.5 RELATIVITY OF REFERENCE 435

4.6 RELATIVITY OF REASONING 466

4.7 CONCLUSION 498

BIBLIOGRAPHY 506

APPENDIX 1 - THE RESEARCH ON THE COLOR DOMAIN 522

APPENDIX 2 - A SUMMARY OF THE ARGUMENT 561

LIST OF TABLES

Table	Description	Page
2.1.1	Comparison of Terminology Used by Various Theorists	51
2.1.2	Terminology Used in the Present Work	52
A.1.1	Summary of the Color Domain Experiments	556

LIST OF FIGURES

Figure	Description	Page
1.2.1	Language as a Causal Determinant of Thought	16
1.2.2	Language and Thought Identified	17
2.2.1	Bloomfield's Paradigm for the Study of Natural Language	66
2.2.2	Discovery Procedures	71
2.2.3	"Environments" based on the Vocabulary 'a', 'b'	73
2.2.4	Matrix for Representing a Corpus of Data	74
2.2.5	Syntactic Transcription of an Utterance from the Corpus	82
2.2.6	A Simple Example of a Sentence-Syntax	93
2.2.7	Phrase Structure Diagram	94
2.3.1	Katz and Fodor's Dictionary Entries	134
2.3.2	Katz and Fodor's Representation of the Meaning of a Syntactically Complex Unit	138
2.3.3	Chomsky's Organization of a Generative Grammar (1965)	140
2.3.4	Jackendoff's Organization of Linguistic and Cognitive Structures	152
2.4.1	A Sample Coordination Problem	232
2.4.2	Platt's Architecture for a Theory of Understanding	239
2.4.3	Structure of a Davidsonian Grammar	260
3.3.1	Relation of the *Geisteswissenschaften* to the Natural Sciences	279
3.3.1	Two Sciences Explaining the Same Causal Relation	305

3.3.2 Two Sciences Explaining the Same Causal Relation According to the "Strong" Reading of Hume's Principle 307

3.3.3 Two Sciences Explaining the Same Causal Relation According to the "Weak" Reading of Hume's Principle 308

3.4.1 Fodor on the Relation Between Psychology and Physics 321

4.2.1 The Dual Method Alternative 392

4.2.2 The Single Method Alternative 393

4.2.3 The Single Method Alternative (Again) 397

A.1.1 Fishman's Systemization of the Whorfian Hypothesis 524

LIST OF SYMBOLS

Logical Implication	->
Logical Equivalence	≡
Logical Conjunction	&
Logical Negation	~
Logical Disjunction	v
Universal quantifier	(x)
Existential quantifier	(Ex)
Quine's quasi-quotes	⌜ ⌝
Set brackets	{ }
Ordered sets	< >
Element of a set	ϵ
Null set	ø
Set Union	U
Set Intersection	∩
Syntactic Production Rule	->
Syntactic Derivation	=>, =>>

CHAPTER ONE - WHORF'S HYPOTHESIS

1.1 WHORF'S HYPOTHESIS

What is the relation between language and thought? In De Interpretatione, Aristotle writes:

> Now spoken sounds are symbols of affectations in the soul, and written marks symbols of spoken sounds. And just as written marks are not the same for all men, neither are spoken sounds. But what these are in the first place - affectations of the soul - are the same for all...[1]

John Locke writes in a similar vein in his Essay:

> Man, though he have a great variety of thoughts, and such from which others as well as himself might receive profit and delight; yet they are all within his own breast, invisible and hidden from others, nor can of themselves be made to appear. The comfort and advantage of society not being to be had without communication of thoughts, it was necessary that man should find out some external sensible signs, whereof those invisible ideas, which his thoughts are made up of, might be made known to others. For this purpose nothing was so fit, either for plenty or quickness, as those articulate sounds, which with so much ease and variety he found himself able to make. Thus we may conceive how words, which were by nature so well

---

[1] Aristotle, Categories and De Interpretatione, translated by J. L. Ackrill, (Oxford University Press, 1963), 43.

> adapted to their purpose, came to be made use of by men as the signs of their ideas...[2]

Aristotle and Locke have similar views on the relation of language and thought. They hold that thinking is an internal process that takes place independently of any linguistic activity. On occasion we want to express or communicate our thoughts, and for that purpose we use language. On this view language is ancillary to thought, occasionally called upon to provide external expression to what is already there internally. It is a corollary of this view that, individual differences aside, all people think the same way, no matter what language they speak.

This conception of the relation between language and thought was challenged by the American linguist, Benjamin Lee Whorf (1897-1941). Whorf questioned Aristotle's easy assumption that the "affectations of the soul" are the same for all men.

> The ethnologist engaged in studying a living primitive culture must often have wondered: 'What do these people think? How do they think? Are their intellectual and rational processes akin to ours or radically different?'[3]

Whorf answered these rhetorical questions with a theoretical approach radically opposed to Aristotle's and Locke's. He claimed that there are profound differences in the way that certain groups think, and that these differences are primarily attributable to the languages spoken by the various groups. The following passage, probably the most famous in Whorf's work, summarizes his view:

---

[2] Locke, J., An Essay Concerning Human Understanding, collated and annotated by A. C. Fraser, (Dover Publishing, 1957), volume II, 8-9.

[3] Whorf, B. L., Language, Thought and Reality, edited by J. B. Carroll, (The M.I.T. Press, 1956), 65.

> When linguists became able to examine critically and scientifically a large number of languages of widely different patterns, their base of reference was expanded; they experienced an interruption of phenomena hitherto held universal, and a whole new order of significances came into their ken. It was found that the background linguistic system (in other words, the grammar) of each language is not merely a reproducing instrument for voicing ideas but rather is itself the shaper of ideas, the program and guide for the individual's mental activity, for his analysis of impressions, for his synthesis of his mental stock in trade. Formulation of ideas is not an independent process, strictly rational in the old sense, but is part of a particular grammar, and differs, from slightly to greatly, between different grammars. We dissect nature along lines laid down by our native languages. The categories and types that we isolate from the world of phenomena we do not find there because they stare every observer in the face; on the contrary, the world is presented in a kaleidoscopic flux of impressions which has to be organized by our minds -- and this means largely by the linguistic systems in our minds. We cut nature up, organize it into concepts, and ascribe significances as we do, largely because we are parties to an agreement to organize it in this way -- an agreement that holds throughout our speech community and is codified into the patterns of our language. The agreement is, of course, an implicit and unstated one, BUT ITS TERMS ARE ABSOLUTELY OBLIGATORY; we cannot talk at all except by subscribing to the organization and classification of data which the agreement decrees.
>
> This fact is very significant for modern science, for it means that no individual is free to describe nature with absolute impartiality, but is constrained to certain modes of interpretation even when he thinks himself most free. The person most nearly free in such respects will be a linguist familiar with many widely different linguistic systems. As yet no linguist is in such a position. We are thus introduced to a new principle of relativity, which holds that all observers are not led by the same physical evidence to the same picture of the universe, unless their linguistic backgrounds are similar, or can in some way be calibrated.[4]

---

[4] *Ibid.*, 212-214.

This passage provides us with a excellent summary of Whorf's Linguistic Relativity Hypothesis. It is the purpose of the present work to evaluate the Linguistic Relativity Hypothesis. I will conclude by rejecting the hypothesis, not because I think it is false on straight-forward empirical grounds, but for reasons more properly labelled "conceptual."

These results lie in the chapters ahead. For now, our task is to attempt a sympathetic understanding of Whorf. I will present a summary of what I believe to be the essential steps in Whorf's argument, but preliminary to that, here are three miscellaneous comments that may help orient the reader.

1. Whorf was not the first to advance the Linguistic Relativity Hypothesis. The German philosopher and philologist, Wilhelm von Humboldt (1796-1835) is often credited with being the originator of the idea.[5] Whorf himself credits a French grammarian, Fabre d'Olivet (1768-1825) as being the real originator[6], and the Irish linguist James Byrnes (1820-1897) as being the next significant advocate.[7] These early statements stood outside the main intellectual currents of their time. It was not until the twentieth century that linguistic relativity became a serious issue for a significant number of thinkers. The

---

[5] B. B. Lloyd, Perception and Cognition: A Cross-Cultural Perspective, (Penguin Books, 1972), 36.

[6] Whorf, B. L., Language, Thought and Reality, edited by J. B. Carroll, (The M.I.T. Press, 1956), 74.

[7] Ibid., 76.

writings of Ernst Cassirer (1874-1945)[8], Franz Boas (1858-1942)[9] and Edward Sapir (1884-1939)[10] did much to promote the thesis. (The Linguistic Relativity Hypothesis is often known as the "Sapir-Whorf Hypothesis.")

The history of the Linguistic Relativity Hypothesis is not the primary concern of the present work. I want to concentrate on Whorf because his version of the Linguistic Relativity Hypothesis is perhaps the most influential in the literature, but more importantly, because I believe that Whorf makes a number of mistakes that are interesting in their own right. I believe that we can profit from a careful dissection of Whorf's errors.

2. Not all cognitive relativists attribute cognitive variation to natural language. That is, there are thinkers, for example the philosopher of science, Thomas Kuhn, who hold that individuals can have radically different world-views even though they share the same natural language.[11] I would venture to guess that the majority of contemporary cognitive relativists are relativists for reasons other than Whorf's; that is, natural language is not the essential factor for most of them.

---

[8] See especially E. Cassirer, The Philosophy of Symbolic Forms (Yale University Press, 1955, 1957) and E. Cassirer, An Essay on Man, (Yale University Press, 1962).

[9] Franz Boaz, "Introduction to the Handbook of American Indian Languages", reprinted in Language, Culture and Society, edited by B. G. Blount, (Winthrop Publishers, 1974).

[10] E. Sapir, Selected Writings of Edward Sapir in Language, Personality and Culture, edited by D. G. Mandelbaum, (University of California Press, 1949).

[11] T. S. Kuhn, The Structure of Scientific Revolutions, 2nd edition, (The University of Chicago Press, 1970).

3. Readers sensitive to the political undertones of the social sciences may be worried that there is a xenophobic element in Whorf's work. Two comments are in order: (a) Moral and political evaluations are completely absent from Whorf's writings. Furthermore, although he stresses the differences in intellectual patterns between members of different cultures, he nowhere suggests that any local intellectual pattern is in any way superior or inferior to any other. The only intellectual pattern that Whorf considers superior is the one which strives to understand and "calibrate" the various local patterns.

(b) Cognitive relativists who do explicitly advance moral and political admonitions almost always do so from a left-wing perspective. That is, some left-wing thinkers embrace relativism because they object to the idea that Western science is the highest form of thought. They feel that this glorification of Western science is a chauvinistic and intolerant attitude, and is linked to other forms of repression. Feyerabend[12] and Barnes and Bloor[13] are examples of this marrying of cognitive relativism and left-wing ideology.

---

[12] P. K. Feyerabend, "How to be a Good Empiricist - A Plea for Tolerance in Matters Epistemological", Philosophy of Science, The Deleware Seminar, edited by B. Baumrin, (The University of Deleware, 1963).

[13] B. Barnes and D. Bloor, "Relativism, Rationalism and the Sociology of Knowledge", Rationality and Relativism, edited by M. Hollis and S. Lukes, (The M.I.T. Press, 1982).

1.2 WHORF'S ARGUMENT

Whorf's arguments are sketchy. Controversial ideas are often presented without the justification they require. Whorf can be partially excused for this, since many of his papers were published posthumously and he may not have considered them finished. On the other hand, it must be recognized that Whorf was never an especially deep or subtle thinker. His ideas are interesting, but not because he provides tight arguments in their favor. Ultimately Whorf's ideas present a challenge: What is the basis of their appeal? John Carroll speculates as follows:

> One wonders, indeed, what makes the notion of linguistic relativity so fascinating even to the nonspecialist. Perhaps it is the suggestion that all one's life one has been tricked, all unaware, by the structure of language into a certain way of perceiving reality, with the implication that awareness of this trickery will enable one to see the world with fresh insight.[14]

I think that Carroll is perhaps correct in this speculation, and consequently the student of the Linguistic Relativity Hypothesis is cautioned against harboring any non-intellectual motivation ("personal growth", "expanded consciousness", etc.) that might make one want the hypothesis to be true.

Whorf's argument for the Linguistic Relativity Hypothesis is spread over several papers written between 1936 and 1941. This argument can be reconstructed as having the following five theses:

---

[14] J. B. Carroll, "Introduction", in B. L. Whorf, Language, Thought and Reality, edited by J. B. Carroll, (The M.I.T. Press, 1956), 27.

1. To think is to employ concepts.
2. Languages have a conceptual structure.
   (That is, a system of concepts is associated with each natural language just as a syntax is associated with each.)
3. Languages differ.
   (That is, different natural languages have different conceptual structures.)
4. Language determines thought.
   (That is, when we learn a natural language we acquire an unconscious command over its conceptual structure, and this conceptual structure becomes a foundation for all our concepts.)
5. Cross-cultural understanding is difficult/impossible.
   (That is, if language A and language B are very different, the concepts of the speakers of language A will be very difficult, if not impossible, to express in language B.)

Each of these five theses will now be explained in greater detail.

1. To think is to employ concepts.

Following Carl Jung, Whorf held that "psychic functions" come in two varieties, the rational or intellectual functions, and the non-rational. The reception of sensations from our sense organs is the primary non-rational function. Whorf held that sensation is the same for all people and that it is not subject to modification by language. In themselves sensations are unstructured; they are a "kaleidoscopic flux" that require organization in order to yield coherent experience.

The organization of sensation is provided by the intellectual function of "thought". (The other intellectual function is "feeling", thus making Whorf's approach compatible with belief-desire approaches to human action.) The essence of "thought" is the application of concepts. Through the application of a system of concepts, we organize our sensations into a structured view of nature.

Another key element of Whorf's conception of thought is the claim that certain concepts are more important than others. For example, Whorf says that the basic concepts held by a speaker of English are space, time, and the concept of the enduring physical object. These "cosmic forms", as he calls them, form a framework in which all our other concepts must find their place.

2. Languages have a conceptual structure.

In claiming that each natural language has a conceptual structure, Whorf did not mean merely that we can express various concepts using the individual terms of a language. Whorf went beyond this and claimed that the various grammatical classes and categories have conceptual correlates. For example, it might be held that in English the grammatical class of nouns is associated with the concept of an object or thing. The grammatical class of verbs is associated with the concept of an activity or process that objects perform or undergo.

Although Whorf would no doubt object to the crudity of my noun/verb example, it will suffice to make the point. The idea is that the concepts associated with nouns and verbs are both deeply embedded into the English language. To use English is to view the world as made up of objects that possess various properties.

It is profitable to examine Whorf's claim against a historical background. Whorf was familiar with two opposing methods used to define and identify grammatical categories. One method is the traditional "notional" approach[15] in which grammatical categories are defined in terms of their conceptual role. This tradition goes back to the Greeks and still persists in the training we give children and in many sophisticated linguistic theories. A prototypical example of the notional approach would be the definition of a noun as any word standing for a person, place or thing.

The other approach is the "formal" or "distributional" approach[16] associated with structuralist linguistics, the dominant linguistic school at the time that Whorf wrote. Practioners of this school completely rejected linguistic concepts that were defined notionally. It was held that notional definitions were based on vague intuitions, and they relied on fuzzy mentalistic or metaphysical terminology. The proper approach, according to the structuralists, was to provide precise operational definitions of linguistic concepts. Specifically, linguistic concepts were to be defined, ultimately, as constructions from a corpus of phonological data. A grammatical category or class, on this view, had to be defined as a set of phonological patterns that stand in a certain relation to other phonological patterns. This was all to be done through the application of an objective method that could be used by any trained person to generate identical results.

---

[15] For a fuller account of the "notional" approach, see J. Lyons, Introduction to Theoretical Linguistics, (Cambridge University Press, 1968), 134.

[16] For a fuller account of the "distributional" approach, see ibid., 143. The distributional approach is discussed in detail in chapter two, section 2.2 of this work. In what follows I generally use the term 'the Bloomfieldian program' to refer to this approach.

Whorf accepted that a formal method of this sort was necessary in order to generate objective results in empirical syntactic studies. However, he felt that once the method has been employed to generate the grammatical categories, it is then necessary to apply a traditional notional definition (he called them "functional" definitions) to each category which would serve to identify its semantic role. In sum, Whorf's view was that neither the traditional notional approach or the then current structuralist "formal" approach is sufficient for defining grammatical categories. Both have to be used to generate a comprehensive account.

> At the beginning of investigation of a language, the "functional" [i.e., notional] type of definition, e.g. that a word of a certain class, say a "noun", is "a word which does so-and-so," is to be avoided when this is the ONLY test of distinction applied; for people's conceptions of what a given word "does" in an unfamiliar language may be as diverse as their own native languages, linguistic educations and philosophical predilections. The categories studied in a grammar are those recognizable through facts of a configurational sort, and these facts are the same for all observers. Yet I do not share the complete

> distrust of functional definitions which a few modern grammarians seem to show. After categories have been outlined according to configurational facts, it may be desirable to employ functional or operational symbolism as the investigation proceeds. Linked with configurative data, operational descriptions become valid as possible ways of stating the MEANING of the forms, "meaning" in such cases being a characterization which succinctly accounts for all the semantic and configurational facts, known or predictable.[17]

Whorf went on to distinguish two types of grammatical categories, which he called phenotypes and cryptotypes. Phenotypes are grammatical categories whose members can be identified entirely on the basis of non-semantic formal patterns. That is, the members will have formal "marks", to use the Post-Bloomfieldian term, in every context in which they appear. Cryptotypes, on the other hand, have members which only show "marks" in some contexts, not all. Whorf gives the example of gender in English, in which all the personal names, species of animals, and many geographical entities and artifacts are assigned to either the masculine or feminine class. Whorf demonstrated that in the English language, gender is manifested in some contexts and not others, thus making it a "cryptic" category from the point of view of a linguist who is trying to specify the grammar of the language. Whorf held that both phenotypes and cryptotypes have meanings or concepts associated with them. Finally, Whorf noted that languages vary in terms of the proportion of cryptotypes to phenotypes. In some languages there are almost no cryptotypes; others, like Hopi, are rich in them.

---

[17] B. L. Whorf, Language, Thought and Reality, edited by J. B. Carroll, (The M.I.T. Press, 1956), 88.

In summary, Whorf held that each natural language has a set of grammatical categories, and associated with each grammatical category there is a "meaning" or concept.

3. Languages differ.

Whorf believed that natural languages can differ profoundly. His views were consonant with those of most other American linguists of his day, and opposed to the views of the "traditional grammarians", i.e., those who believed that the notionally defined grammatical categories of European languages (especially Latin) are applicable to all natural languages. American linguists of the early twentieth century were deeply involved in a major collective project, that of recording and analyzing the American Indian languages. It is not surprising that they rejected the assumptions of the traditionalists, and insisted that each language be studied on its own terms.

However, as pointed out above, Whorf differed from many of his contemporaries in his willingness to go beyond formal syntactic categorization and to associate concepts with the categories. In other words, Whorf insisted that languages differed not only on a formal configurational basis, but on the level of meanings as well.

4. Language determines thought.

The next step of Whorf's argument is the claim that the concepts we employ in thinking actually derive from language. Most significantly, the concepts associated with the grammatical categories of our native language become the central concepts of our overall conceptual scheme. Moreover, the concepts associated with cryptotypes tend to be more fundamental, in terms of our resulting conceptual scheme, than those associated with phenotypes. In the following passage Whorf even suggests a hypothesis regarding the historical genesis of concepts; they arise when a phenotype undergoes a linguistic transformation into a cryptotype:

> As outward marks become few, the [grammatical] class tends to crystallize around an idea - to become more dependent on whatever synthetizing principle there may be in the meanings of its members. It may be even be true that many abstract ideas arise in this way; some rather formal and not very meaningful linguistic group, marked by some overt feature, may happen to coincide very roughly with some concatenation of phenomena in such a way as to suggest a rationalization of this parallelism. In the course of phonetic change, the distinguishing mark, ending, or what not is lost and the class passes from a formal to a semantic one... As time and use goes on, it becomes increasingly organized around a rationale, it attracts semantically suitable words and loses former members that now are semantically inappropriate... Semantically it has become a deep persuasion of a principle behind phenomena, like the ideas of inanimation, of "substance", of abstract sex, of abstract personality, of force, of causation - not the overt concept (lexation) corresponding to the word causation, but the covert idea, the "sensing", or, as it is often called, ...the "feeling" that there must be a principle of causation. Later this covert idea may be more or less duplicated in a word and a lexical concept invented by a philosopher: e.g., CAUSATION.[18]

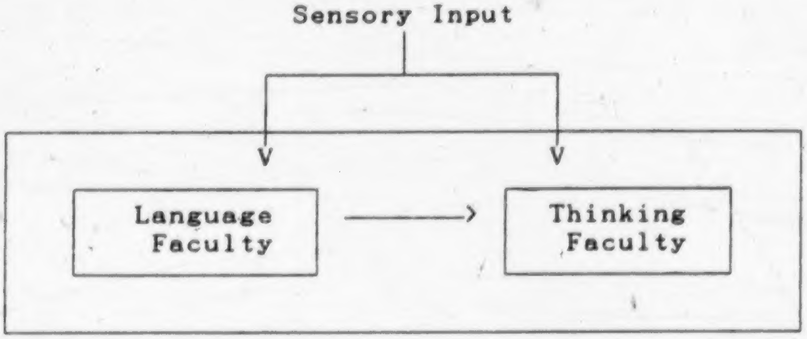
Now for the disappointing part. Whorf does not give us a clear picture of how language is supposed to determine the thought of an individual language user. One would expect that Whorf would provide a theory of linguistic/cognitive development to explain how the child's thought is shaped as he acquires language. In fact, Whorf does not even clearly differentiate two fundamentally different ways of interpreting the claim that language determines the thought of an individual speaker. One interpretation of this claim is that there is a causal relation between language and thought. The other - non-causal - interpretation is that thinking essentially involves the use of a natural language, and that consequently as languages differ so will thought. Roughly speaking, the latter interpretation entails that thinking is nothing other than internalized language use. On this view language and thought are not two different things, and therefore no causal relation can exist between them.

---

[18] Ibid., 80-81.

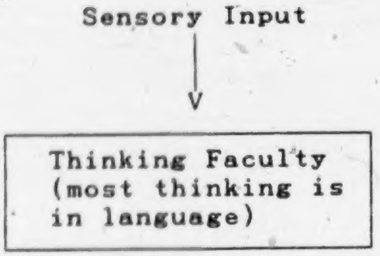
One way of making these contrasting interpretations more precise is to construe them as alternative theories about psychological faculties; that is, alternative theories about the number of faculties and the relations between them. On the first interpretation, then, there are two distinct psychological faculties, one for language use and one for thinking. Children learn their native language, which is represented in the language faculty. This representation causes a similar conceptual structure to develop in the thinking faculty, and consequently as observations of the external world are received into the thinking faculty, they are assimilated into that conceptual framework.

Figure 1.2.1
Language as a Causal Determinant of Thought



In the second interpretation there is only one faculty which combines the functions of language and thought. On this view, to think is to use language. As children acquire language their ability to think grows in proportion. Also, the style in which they think will depend on the grammatical categories of the language that they learn. There is no causal relation postulated in this theory, for there is no pair of objects to postulate a causal relation between.

Figure 1.2.2
Language and Thought Identified



It is worthy of note that each interpretation sets its own limits on the Linguistic Relativity Hypothesis. Take the causal interpretation. It states that the development of the thought faculty is causally influenced by the language faculty. However, for the language faculty to develop, the child would, at a minimum, already have to be able to recognize the existence of a populated world where people engage in linguistic communication. That is, the development of the language faculty seems to presuppose that the child already has a lot of very fundamental concepts and consequently these concepts will be part of everyone's conceptual scheme, regardless of the language learned.

On the other hand, the non-causal interpretation seems to lessen the impact of the Linguistic Relativity Hypothesis precisely because causality is lacking. It seems to suggest that since people think in natural languages, the appropriate method for comparing the thoughts of speakers of different languages would be to translate sentences expressing the thoughts from one speaker's language to the other's. If the translation is straightforward, then the speakers think alike. If translation is difficult (or perhaps even impossible), then the speakers think very differently. But note what is happening here. We are shifting from a substantive issue about how these people really think, to the (controversial) methodological issue of how we should go about translating them. This slide from substance to method seems difficult to avoid in the non-causal interpretation.

One is hard pressed to find a passage in Whorf that definitively settles which interpretation is the correct one. On balance though, various passages suggest that Whorf intended the non-causal interpretation.[19] On the other hand, Whorf seems to intend the Linguistic Relativity Hypothesis to be a substantive claim about human psychology. This sets up a tension in Whorf's work to which I will return in later chapters. Indeed the contrast between the causal and non-causal (or, alternatively, the substantive and the methodological) interpretations will pervade chapter four.

5. Cross-cultural understanding is difficult/impossible.

Whorf claimed that if two languages are very different in terms of grammatical structure and the concepts associated with grammatical categories, then it will be very difficult to express in one language what can be said easily in the other. Whorf may have intended something even stronger: namely that cross-cultural understanding is impossible in cases of extreme differences between languages. Unfortunately, Whorf's writings are not clear on this issue. Consider the following four points:

---

[19] See especially *ibid.*, 65-69.

(a) Whorf does say that all languages are "observationally adequate" by which he means that all languages are capable of fully accounting for all the possible sensory input that the world may cause us to have.[20] At this point Whorf could have easily followed the course of many philosophers of his time and adopted a foundationalist approach to semantics (in which it is hoped that the meanings of expressions can be built up from a foundation of elementary terms that have "sensory meanings"), which, on the face of it, seems to hold out the promise of providing a means of comparing languages: that is, any conceptual differences between languages could be ultimately reconciled at the level of "sensory meanings." However, Whorf did not take this route, which is just as well, given the problems associated with foundationalist semantics.[21]

(b) Whorf claims that a comparative linguist who has studied, through careful objective methods, a great number of languages, is in a privileged position. Such a linguist would be able to "calibrate" the various languages[22], but the exact meaning of "calibration" is not clear. In any case, Whorf also suggested that comparative linguistics has not yet progressed to the point where such "calibrations" are possible.

---

[20] Ibid., 58.

[21] See W. V. O. Quine, "Two Dogmas of Empiricism", From a Logical Point of View, (Harper and Row, 1961), and J. Dancy, Introduction to Contemporary Epistemology, (Blackwell, 1985), part II.

[22] B. L. Whorf, Language, Thought and Reality, edited by J. B. Carroll, (The M.I.T. Press, 1956), 214.

(c) Whorf said that the conceptual system of the Hopi Indians is properly expressible only in the Hopi language,[23] which does seem to suggest incommensurability, and thus contradicts point (b).

(d) Whorf did not discuss the relevance of bilingualism to this issue. Does the bilingual have one conceptual scheme, two that are internally "calibrated", or two that are not "calibrated"? Whorf offers no answers to these questions.

As points (a) - (d) indicate, Whorf did not present a clear account of how far we can go in comparing languages and conceptual schemes. For now, let us simply note that a distinction can be made between a "strong" version of the Linguistic Relativity Hypothesis, according to which at least some pairs of conceptual schemes are incommensurable, and a "weak" version, according to which at least some degree of comparison is possible between all conceptual schemes, even where they differ significantly. This distinction will be raised again in Chapter Four.

---

[23] Ibid., 58.

1.3 THE WESTERN MIND AND THE HOPI MIND

The reconstruction of Whorf's argument in the previous section will now be supplemented with a presentation of two specific examples of the effect of language on thought. These two examples are the only two that Whorf developed in his own work. One is an analysis of what Whorf called Standard Average European (SAE) and its effect on the Western mind. Whorf felt that the European languages are similar enough in grammar as to have a uniform effect on the thinking of their speakers. The contrasting example is the language of the Hopi Indians. These examples will be presented as Whorf presented them, with no comment on their plausibility.

According to Whorf, the Western world has imposed two grand "cosmic forms" upon the universe, and these forms provide us with our basic orientation to the observable world. These forms are time and space. Since Kant, these forms have often been interpreted as being universal forms of human sensibility. Whorf objects to this view:

> It is sometimes said that Newtonian space, time and matter are sensed by everyone intuitively, whereas [Einstein's] relativity is cited as showing how mathematical analysis can show how intuition is wrong...[L]aying the blame on intuition for our slowness in discovering the mysteries of the cosmos, such as relativity, is...wrong...The right answer is: Newtonian space, time and matter are no intuitions. They are recepts from culture and language. That is where Newton got them.[24]

---

[24] B. L. Whorf, Language, Thought and Reality, edited by J. B. Carroll, (The M.I.T. Press, 1956), 152-153.

Time and space, then, are not universal features of human experience; they are the resultant of the particular structure of SAE. And likewise, the physics that is built on these concepts is not the only way that physical science could be organized. It is possible, according to Whorf, to construct an entirely different physics without the concepts of space and time, and this alternative physics could provide an account of natural phenomena that is just as coherent and exhaustive as Newton's.

What are the linguistic features of SAE responsible for the concepts of time and space? Whorf claims that the prelinguistic experience of time is rather unstructured; the most that we could experience would be a consciousness of "becoming later-and-later". However, the SAE languages add structure to these experiences by "objectifiying" words like 'summer', 'winter', 'September', 'morning', 'noon', and 'sunset'. By 'objectification', Whorf means that the SAE languages treat these terms in the same way that they treat terms that refer to physical objects; consequently we can say "at sunset" and "in winter" just as easily as we can say "at the corner" or "in the orchard".[25] The phenomenon of objectification is further reflected in the fact that in SAE we can count "non-actual" (Whorf's terminology) entities. Whorf states that an entity is actual only if it can be completely and directly present in experience. A summer, then, is a non-actual entity. The fact that summers can be counted is further evidence that our language objectifies time. Other languages do not incorporate this objectification mechanism; they produce less distortion to the basic pre-linguistic experience of time. Where we would say "a summer" and "summers", other languages would say something like "summering". A consequence of our objectification of time is that we see it in terms of a spatial analogy, as a number of entities strung out in one dimension, one after another. This notion of time, fundamental to Newtonian physics, is completely foreign to the speakers of certain non-SAE languages.

---

[25] Ibid., 142.

Space, on the other hand, is not so subject to the constructive influence of language. Whorf felt that the human mechanisms of visual perception provide a pre-linguistic experience of space which is much more well organized than our pre-linguistic experience of time. However, languages can differ in the degree to which they call attention to the notion of physical space, and consequently, speakers of various languages can become differentially attuned to the importance of physical space as a principle of cognitive organization. SAE does, in fact, call great attention to physical space. This is accomplished, once again, by the linguistic phenomenon of objectification, that is, by treating many terms that do not refer to physical objects according to the same linguistic conventions that are employed for terms that do refer to physical objects. In SAE, physical space and the objects that it contains are linguistic exemplars, according to which other entities are also treated. Physical space becomes a metaphor for many things that are essentially non-spatial. The overall effect of this pervasive metaphor is that physical space is elevated to a very important principle in the cognitive organization of the speakers of SAE. Space is elevated to the status of a cosmic form.

In terms of the syntax of SAE, the objectification phenomenon can be traced to a number of grammatical categories, both phenotypes and cryptotypes. Whorf mentions the treatment of plurality and cardinality in SAE, along with the existence of mass nouns, the three tense verb system and the significant number of terms expressing duration, intensity, and tendency.[26] The net effect is that the SAE speaker, then, lives in a world organized around the cosmic forms of space and time.

The Hopi language, on the other hand, does not treat terms referring to non-physical entities the same way as those terms that refer to physical objects. Whorf comments that it is as if there were a taboo against objectification in the Hopi language. Consequently, in the Hopi world view the concepts of space and time are replaced with two other cosmic forms. This pair of concepts play a central role on Hopi cognitive organization, just as space and time do for speakers of SAE. These cosmic forms are called "manifested" and "manifesting" by Whorf, but he also uses the alternative terms "objective" and "subjective".

Roughly speaking, the first form consists of all that has been physically actualized, while the second form consists of all that is not. The second category would include all that is conveyed by the English terms "future", "mental", "possible" and "unrealized." However, there is more to the second form than just this; there is also an element of dynamism or force. This force relates the two forms in an overall metaphysics; the force produces a constant tendency for the unmanifested to become manifest, for the subjective to be become objective. This tendency or process results in there being something like a "border" between the two realms; that is, there is a point of actualization, where what was previously subjective becomes objective. According to Whorf, experience is mostly objective, but experience also includes the "border" of newly emerging manifestation.

---

[26] Ibid., 139-147.

An interesting consequence of the Hopi cosmic forms, according to Whorf, is that the Hopi have no conception of what we refer to as the temporal future. Because of the objectification inherent in SAE, we tend to think of the past and the future as different parts of the same continuum. But for the Hopi, such an identification is impossible. In the Hopi world view the manifested realm corresponds roughly to an undifferentiated combination of our past and present physical space. The realm of the manifesting, on the other hand, corresponds to an equally undifferentiated amalgam of what we think of as the future, the mental, the possible, etc., all of which are quite distinct from the physical. Consequently the Hopi cannot think of the non-actualized realm in terms of a distance metaphor that is derived from the physical world. It is not possible, according to Whorf, that the Hopi could conceive of future time as being measurable. Furthermore, as the following passage indicates, the Hopi tend to run space and time together even when they are considering the manifested realm:

> What happens at a distant village, if actual (objective) and not a conjecture (subjective) can be known 'here' only later. It does not happen 'at this place', it does not happen 'at this time'; it happens at 'that' place and 'that' time. Both the 'here' happening and the 'there' happening are in the objective, corresponding in general to our past, but the 'there' happening is the more objectively distant, meaning from our standpoint that it is further away in the past just as it is further away from us in space as the 'here' happening.[27]

How does Whorf explain the linguistic origins of these exotic world views? He traces it primarily to certain categories of Hopi verbs, specifically to the variations in "aspect" that the Hopi language observes. "Aspect" is a linguistic term which refers to different forms that a verb can take in order to represent different "phases" of the action or state described by the verb. Different aspects can indicate whether the action or state is completed or in progress, as instantaneous or enduring, as momentary or habitual, etc. Whorf has identified nine aspects in Hopi,[28] but it is the pervasive influence of just two of these aspects which has such important consequences for the world view of the Hopi. One of these is the "inceptive", or as it is alternatively called, the "inchoative"; an aspect which expresses an action commencing or "being started." According to Whorf, the use of the inceptive is related to the cosmic form of the manifested, because the inceptive draws attention to the "border" between manifested and manifesting from the point of view of the manifested side. The other relevant "aspect" is the "expective." This term, by the way, is Whorf's own. The expective brings attention to the border between the two forms from the manifesting side. Together, these two aspects of Hopi verbs have as decisive an effect on the speakers of Hopi as the objectification feature of SAE has on its speakers.

---

[27] Ibid., 146.

[28] Ibid., 51.

Whorf's discussion of the Western mind and the Hopi mind invites commentary from many angles. I will hold my comments for Chapters Two and Three, where I discuss language and thought in more detail.

1.4 MIND: CONTENT OR CONSCIOUSNESS?

Whorf's thesis is that language affects thought by determining the central concepts that are used when one thinks. In expressing the Linguistic Relativity Hypothesis this way, Whorf is implicitly aligning himself with a particular approach to the philosophy of mind, and he is divorcing himself from an alternative approach. In this section I want to briefly address these two approaches in relation to the Linguistic Relativity Hypothesis.

In the period of modern philosophy, dating from the early seventeenth century, two distinct approaches to the philosophy of mind have emerged. These two approaches can be characterized as alternative statements of what mentality is; that is, they can be characterized as two attempts to specify a criterion of the mental.

One of the two approaches focuses on consciousness or experience as the defining characteristic of the mental. Since Descartes, this approach has been linked to epistemological doctrines. That is, advocates of this approach have often claimed that each of us has a special epistemological relation to our mental states, such that we could never be mistaken or wrong in making a sincere assertion about the mental states we are undergoing. This has been called the "incorrigibility criterion" of the mental.[29] The epistemological doctrines associated with the incorrigibility criterion have been subject to frequent attacks[30] and are currently held in low repute. Similarly, the "consciousness" approach has been historically linked with the dubious ontological claim that minds are non-physical entities (Descartes' philosophy is a prime example), which again, is held in low repute. The "consciousness" approach, then, has kept some bad company, with the result that some twentieth century thinkers[31] have concluded that most of what we think about "consciousness" is hopelessly confused, that it is some sort of cultural aberration that was kicked off by Descartes and confounded by many others.

The other approach focuses on the possession of "content" as the defining mark of the mental. This approach was articulated by Franz Brentano, who reintroduced the term 'intentionality' to characterize the possession of content:

---

[29] K. V. Wilkes, Physicalism, (Humanities Press, 1978), 4.

[30] For example, J. Dancy, Introduction to Contemporary Epistemology, (Blackwell, 1985), chapter 4.

[31] For examples see G. Ryle, The Concept of Mind, (Penguin University Books, 1949), and R. Rorty, Philosophy and the Mirror of Nature, (Princeton University Press, 1949).

> Every mental phenomenon is characterized by what the scholastics called the intentional (and also mental) inexistence of an object, and what we would call, although not in entirely unambiguous terms, the reference to a content, a direction upon an object..., or an immanent objectivity. Each one includes something as an object within itself, although not always in the same way. In presentation something is presented, in judgment something is affirmed or denied, in loved [something is] loved, in hate [something is] hated, in desire something is desired, etc....
>
> This intentional inexistence is exclusively characteristic of mental phenomena. No physical phenomenon manifests anything similar. Consequently, we can define mental phenomena by saying that they are such phenomena as include an object within themselves...[32]

These two approaches have their own histories, and they stand in an uneasy relation to one another. The consciousness approach seems to better accommodate sentient states such as pain, emotion, and the experience of seeing a red object. The content approach seems better suited for sapient states such as applying a concept, holding a belief, and desiring something. Many writers have concluded that the approaches are complementary, that both are needed to give a complete account of the mental. The following passage exemplifies the complementarity thesis:

> ...Intentionality is not the same as consciousness. Many conscious states are not Intentional, e.g., a sudden sense of elation, and many Intentional states are not conscious, e.g., I have many beliefs that I am not thinking about now and may never

---

[32] From F. Brentano, Psychologie vom Empirischen Standpunkt, 1874, quoted in R. Chisholm, "Intentionality", The Encyclopedia of Philosophy, edited by P. Edwards, (MacMillan Publishers, 1967), volume 4, 201.

> have thought of. For example I may believe that my paternal grandfather spent his entire life within the continental United States but until this moment I never consciously formulated or considered that belief. ...[T]he class of conscious states and the class of Intentional states overlap but they are not identical, nor is one included in the other.[33]

Whorf's formulation of the Linguistic Relativity Hypothesis is based on a content approach to mind. Whorf does not make any reference to Brentano, nor does he indicate an appreciation of the subtler problems associated with the content approach. All the same, with his heavy reliance on the "concept" concept, he clearly belongs in the intentionalist camp.

In the chapters that follow, I will argue against the Linguistic Relativity Hypothesis. I will be using facts about intentionality and about the relation between intentional ascriptions and linguistic ascriptions to make this point. My arguments are intended to show that language and mind are not really two different things or processes. Instead, I will claim that when someone interprets another's behavior as the action of a person, the interpreter simultaneously makes linguistic and intentional attributions that are conceptually linked. I will then use facts about the methodology of translation and interpretation to conclude that all persons must share substantially the same world view if translation and interpretation is to work at all. However, these arguments (assuming that they are sound) are effective only against the Linguistic Relativity Hypothesis expressed as a thesis about the effect of language on intentional states. These arguments would not necessarily satisfy someone who feels that mentality is primarily a matter of sentience and consciousness rather than sapience and content.

---

[33] J. R. Searle, _Intentionality_, (Cambridge University Press, 1983), 2-3.

I want to briefly address a challenge from an imaginary psycholinguist that I will call R.W. (Rene Whorf). Like Whorf, R.W. believes that language affects the mind, but R.W. has a very different approach to "mind". R.W. feels that language influences the actual conscious experience of the language user. R.W. is not interested in the mild claim that Hopi concepts are different from the concepts of English speakers. Rather, he is making the much stronger and livelier claim that what it is like to be a Hopi is very different than what it is like to be a speaker of English. R.W. is familiar with the line of argument that I will be presenting in later chapters, but he is unimpressed. He says:

> Your argument gets off the ground only because you have fooled yourself into thinking that mentality is merely a matter of third-person attributions. That is, in your opinion person A has a mental state only if there is some other (actual or possible) person B such that B is interpreting A's behavior by making linguistic and psychological attributions about A. It is very easy for you (and all intentionalists) to slide into an instrumentalist attitude about mental states, that they are merely convenient constructs for interpreting behavior. And then it is very easy for you to turn your discussion of the Linguistic Relativity Hypothesis into a study of the methodology of third-person interpretation.
>
> But don't you see that this is wrong? Those people whom you are interpreting are not just mindless automata with complex behaviors. They also have conscious experiences, and these conscious experiences are not addressed at all by the third-person approach that you have allowed yourself to slide into. No matter what conclusions you come up with about the

> methodology of interpreting behavior, they are irrelevant to my version of the Linguistic Relativity Hypothesis. For I hypothesize that the conscious experience of the Hopi differs significantly from the conscious experience of the English speaker. Whether I am right or not cannot be settled by an analysis of the methodology of the intentional interpretation of behavior. Whether I am right or not is determined only by certain objective facts: viz., whether or not the conscious experience of the Hopi does actually differ significantly from the conscious experience of the English speaker. These facts are what the Linguistic Relativity Hypothesis is really about. The problem with B.L. Whorf's formulation of the Linguistic Relativity Hypothesis, and your response to it, is that you have both found a way to ignore these facts.

I find R.W.'s remarks to be extremely compelling, even though they are highly critical of the approach that I will be taking in later chapters. His comments conform to my intuitions about the reality of the sentient experience of people and other animals. What response, then, can be made to this seemingly powerful critique?

R.W.'s position is based on (a) the intuition that there are facts about each person's conscious experiences and (b) the claim that these facts cannot be reduced to or identified with facts about his behavior or any other part of the material world. Three possible responses, then, are that R.W. is wrong about (a), or that he is wrong about (b), or thirdly, he is correct. These three responses will be examined in turn.

Response 1: Rejection of Intuitions About Consciousness

One way of defending my approach to the Linguistic Relativity Hypothesis is to deny the validity of R.W.'s intuitions about consciousness. A first step is to deny that these intuitions have any privileged status, as Paul Churchland does in the following passage:

> ...we are here denying the near-universal conviction that the mind or self is somehow 'better known' to itself than is the universe around it. ...Moreover, to put the matter thus...is to highlight the possibility of conceptual PROGRESS in the matter of self-comprehension, and to raise the question of the adequacy of our current framework for internal states...On the view here embraced it is conceivable, first, that said framework is inadequate as a representation of our internal reality, perhaps profoundly inadequate, and second, that one might learn to comprehend and report one's internal states within a different and more adequate framework.[34]

Starting from Churchland's position that our intuitions about consciousness and experience are not sacrosanct, some authors have gone on to argue that they are, in fact, extremely misleading and that they should be banished from our conceptual framework. Gilbert Ryle argued this way in his seminal The Concept of Mind (Ryle's famous phrase was "the Myth of the Ghost in the Machine"), and more recently Richard Rorty has done so in his influential Philosophy and the Mirror of Nature. Ryle and Rorty would counsel me to continue with my behaviorally-based intentional approach and to ignore the ill-founded admonitions of R.W.

Unfortunately, I cannot take comfort in Ryle and Rorty's advice. They may turn out to be correct in the long run, but their specific arguments fail to convince me that we can safely turn our backs on our intuitions about consciousness. Through their work and the work of a legion of other like-minded writers, we have learned much about the conceptual traps and errors that are associated with the consciousness approach. In particular, they have exposed the faulty epistemological and metaphysical doctrines that have been historically associated with the approach. But on the other hand, I do not feel that these criticisms add up to grounds for a wholesale rejection of the idea of conscious experience.

---

[34] P. M. Churchland, Scientific Realism and the Plasticity of Mind, (Cambridge University Press, 1979), 99.

The work of Thomas Nagel[35] has done much to fortify my conviction that Ryle and Rorty are giving consciousness a premature burial. Nagel argues that some of the creatures that exist in the objective world have their own subjective points of view, and that a complete accounting of reality must include subjectivity as well as objectivity. Nagel makes his case not by presenting specific epistemological or metaphysical doctrines, but rather, by showing how the subjectivity/objectivity distinction permeates our thinking about a wide range of philosophical problems including the mind/body problem, personal identity, freedom, knowledge, value, ethics, and the meaning of life. He does not pretend to have a grasp of how subjectivity is to be fully incorporated into an overall view of reality, but at the same time he is highly critical of those, like Ryle and Rorty, who dismiss subjectivity as a pseudo-problem, prematurely likening it to the problem of "witches" or the problem of "phlogiston", problems that were swept away by the advance of science. The following passages from Nagel summarize his position:

---

[35] Especially the last four essays in T. Nagel, Mortal Questions, (Cambridge University Press, 1979), and the first six chapters of T. Nagel, The View From Nowhere, (Oxford University Press, 1986).

> There is a persistent temptation to turn philosophy into something less difficult and more shallow than it is. It is an extremely difficult subject, and no exception to the general rule that creative efforts are rarely successful.[36]
>
> What is needed is something we do not have: a theory of conscious organisms as physical systems composed of chemical elements and occupying space, which also have an individual perspective on the world, and in some cases a capacity for self-awareness as well. In some way that we do not now understand, our minds as well as our bodies come into being when these materials are suitably combined and organized. The strange truth seems to be that certain complex, biologically generated physical systems, of which each of us is an example, have rich non-physical properties. An integrated theory of reality must account for this, and I believe that if and when it arrives, probably not for centuries, it will alter our conception of the universe as radically as anything has to date.[37]

It is clear from these passages that Nagel agrees with Churchland that our current intuitions are not the last word; he obviously feels that progress is possible. However, unlike Ryle and Rorty, he does not equate progress with denial.

My conclusion is that I cannot avoid R.W.'s challenge by claiming that his intuitions about consciousness are faulty. Those intuitions are certainly lacking in clarity, and they may turn out to be illusory, but alternatively, they may be the primitive precursors of a Nagelian science of subjectivity.

---

[36] T. Nagel, The View From Nowhere, (Oxford University Press, 1986), 12.

[37] Ibid., 51.

Response 2: Reduction of Consciousness to the Material Realm

Another response to R.W. is to concede the validity of his intuitions, but to argue that despite appearances, consciousness can be reduced to (or shown to be identical with, or supervenient upon) the material realm. There is a great literature on this topic, and as Anthony O'Hear points out in this passage:

> There is a great divide here between those philosophers like Wilkes, Dennett and Rorty, who are prepared to think of sentience largely in terms of the self-monitoring of neurophysiological systems, in terms, in other words, of its capacity to enable these systems to react flexibly to the environment, and those philosophers, like Nagel, Kripke and Searle, who feel that this leaves out of the picture the most important, indeed the essential aspect of sentience, in which, as Kripke puts it, 'its immediate phenomenological quality' is what pain is, whatever its function or underlying physical cause might be.[38]

If consciousness could be accommodated within a physical framework, then the Linguistic Relativity Hypothesis could be studied by the techniques of physical science, presumably by making neurophysiological measurements. However, once again, I agree with Nagel that our current understanding of these issues is in its infancy. A material theory of the mind may turn out to be the correct one, but in order to be adequate, that theory would have to provide a plausible account of the subjectivity of experience. None of the current proposals come anywhere near accomplishing this task. Consequently, it is premature to reject R.W.'s concerns on the grounds that sentience can be accommodated in a physicalist world view.

---

[38] A. O'Hear, What Philosophy Is, (Penguin Books, 1985), 228.

Response 3: R.W.'s Concerns are Valid

The third response is to accept R.W.'s critique as valid. I am not sure that this is the correct thing to do, but I am equally unsure that R.W. can be dismissed. Not knowing how either to reject or accommodate R.W.'s full-blooded version of the Linguistic Relativity Hypothesis, I will restrict myself to the Linguistic Relativity Hypothesis as Whorf stated it: as a thesis about the effect of language on the mind, intentionally understood.

It does not bother me to make this restriction, I see it as a sensible way to respond to the lack of conceptual tools that are available for attacking R.W.'s problem. In making this restriction I am acknowledging that what I write in the following chapters will not be the last word on the topic, but who would have thought otherwise?

1.5 EMPIRICAL RESEARCH

Whorf's writings received considerable attention in the 1950s and 60s, much of it critical of the lack of precision in Whorf's work. One of the most influential papers in this vein was Eric Lenneberg's "Cognition in Ethnolinguistics",[39] in which Lenneberg characterized Whorf's method of determining cognitive facts as the "translation method". According to Lenneberg, Whorf would often provide a translation of an utterance from an exotic language, such as Hopi, and then proceed to compare the odd-sounding English translation with an English sentence that might be uttered by a native English speaker in the same situation. Lenneberg found fault with this method on a number of counts. First, Whorf typically would provide translation of the terms of the exotic utterances, and then present the synthetic meaning of the utterance as simply a string of meanings in the same order as the terms of the utterance. In contemporary terms, Lenneberg showed that Whorf had failed to provide a solution for what Katz and Fodor have called the "projection problem", that is, the problem of showing how the meanings of the elements of an utterance synthetically contribute to the overall meaning of the utterance.[40]

---

[39] E. H. Lenneberg, "Cognition in Ethnolinguistics", Language, 1953, 29, 463-471. Reprinted in Language in Thinking, edited by P. Adams, (Penguin Books, 1972).

Lenneberg also criticized Whorf's method of translation for its lack of recognition of the metaphorical element in language. Whorf's approach is based on the assumption that everything said by the speaker of an exotic language must be taken literally. Eleanor Rosch has demonstrated how even the translation of French on this basis can result in the attribution of unusual beliefs to speakers of that language.[41] Lenneberg concludes, on the basis of these two faulty aspects of Whorf's method of translation, that the cognitive data that Whorf has identified are an artifact of the method, rather than a set of objective facts.

---

[40] J. J. Katz and J. A. Fodor, "The Structure of a Semantic Theory", Language, 1963, 39, 170-210. Reprinted in Readings in the Psychology of Language, edited by L. A. Jakobovits and M. S. Miron, (Prentice-Hall, 1967), 475.

[41] E. Rosch, "Linguistic Relativity", Human Communication: Theoretical Explorations, edited by A. Silverstein, (Lawrence Erlbaum Associates, 1974), 97-98.

Lenneberg then went on to generalize this conclusion, and to claim that translation between languages of "totally different cultures" is always inexact, always only a "very rough approximation of what has been said and intended originally".[42] Consequently, cognitive data adduced from the method will never yield objective facts, and therefore the method must be abandoned in any serious discussion of the Linguistic Relativity Hypothesis. Instead, argued Lenneberg, cognitive facts must be determined independently of linguistic evidence. The researcher must turn to the non-linguistic behaviors that are "indicative of memory, recognition, learning, problem solving, concept formation and perception".[43] In fact, the areas of study that Lenneberg thought were most appropriate for investigating the Linguistic Relativity Hypothesis are just those domains that can be operationalized by the methods of psychophysics, such as color perception. The restriction to domains that have been operationalized by the methods of pyschophysics means that our linguistic concerns are also restricted to what Lenneberg calls the "language of experience"[44] that is, to terminology, such as color terms, that refers to these cognitive domains.

---

[42] E. H. Lenneberg, "Cognition in Ethnolinguistics", Language, 1953, 29, 463-471. Reprinted in Language in Thinking, edited by P. Adams, (Penguin Books, 1972), 161-162.

[43] Ibid., 164.

[44] E. H. Lenneberg and J. M. Roberts, "The Language of Experience: A Study in Methodology", International Journal of American Linguistics, 1956, 22, supplement.

Lenneberg's ideas were widely accepted and reiterated by the psychologists of the day. The result was that the Linguistic Relativity Hypothesis was recast into a number of experimental designs that drastically reduced the original scope of Whorf's ideas. Much of this work concentrated on the so-called "color domain". The idea was to experimentally determine if color vocabulary had an effect on the perception and the ability to remember colored objects. Research on these lines was carried on for several decades, with most (although not all) of the researchers involved concluding that the Linguistic Relativity Hypothesis had been empirically refuted.[45] Eleanor Rosch, a major participant in this research, summarizes this view in the following passage taken from an article in which she reviews this research:

> We began with the notion of linguistic relativity defined in terms of insurmountable differences in the world view of cultures brought about by differences in natural languages. Because of the variety of requirements for specificity and cross-cultural controls in testing such assertions, we were reduced to the far less sweeping claim that color names affect some aspects of thought. However, we discovered that color appears to be a domain suited to demonstrate just the opposite of linguistic relativity, namely, the effect of the human perceptual system in determining natural categories... At present, the Whorfian hypothesis not only does not appear to be empirically true in any major respect, but no longer even seems profoundly and ineffably true.[46]

Rosch is making two claims here. First, she is agreeing with Lenneberg that Whorf's Linguistic Relativity Hypothesis is only worthy of consideration if it is first transformed into specific empirical hypotheses. Secondly, she is claiming that suitably transformed, the Linguistic Relativity Hypothesis has been shown to be false. I have nothing to say about the second claim (other than that Rosch has generalized too hastily from a limited number of cases). My real concern is with the first claim. Lenneberg and Rosch are correct in criticizing Whorf's methods of attributing linguistic and psychological properties to people. However, they are wrong in thinking that Whorf's ideas must be recast as a set of narrow experimental designs. Whorf's original intentions do not survive this transformation. In particular, Whorf's claim that syntactic categories have semantic correlates with cognitive consequences was lost in the shuffle.

---

[45] See Appendix 1 for a review of this literature.

[46] E. Rosch, "Linguistic Relativity", Human Communication: Theoretical Explorations, edited by A. Silverstein, (Lawrence Erlbaum Associates, 1974), 118-119.

Fortunately there is another alternative, which is to seriously pursue Whorf's goal of trying to come up with a method of making linguistic and psychological attributions that are as objective as possible. That is the alternative that I will pursue in the following chapters.

The empirical research that was spawned by Lenneberg's critique of Whorf's Linguistic Relativity Hypothesis is interesting in its own right (e.g. it has generated Rosch's theory of the "internal structure" of perceptual categories[47]). However, this research must be recognized as a wrong turn in the study of the Linguistic Relativity Hypothesis. Whorf's views cannot be reduced to simplistic claims about the effect of color vocabulary on memory. Whorf's views are quite abstract, and have high level theoretical and methodological implications. That is the level at which I will be addressing the Linguistic Relativity Hypothesis.

---

[47] E. Rosch, "On the Internal Structure of Perceptual and Semantic Categories", *Cognitive Development and the Acquisition of Language*, edited by T. Moore, (Academic Press, 1973).

1.6 PREVIEW

This section is a preview of the chapters ahead.

Chapter Two - Natural Language

Whorf believed that syntactic categories have semantic correlates. This idea is at the heart of his Linguistic Relativity Hypothesis. However, his theory of phenotypes and cryptotypes is far from precise, and his overall conception of how "meaning" is to be studied is far from clear. Consequently, this chapter begins with an examination of Bloomfield's linguistic program. This program, which was a reference point for Whorf's work, was a good deal more precise in terms of its goals and assumptions. However, I will criticize Bloomfield's conception of semantics, and argue that there is no obvious way in which it can be modified in order to be satisfy Whorf's theoretical requirements.

Next I examine another linguistic program, which I call "conceptual semantics". This program has elements of Ferdinand de Saussure's "structuralism", and is often combined with Noam Chomsky's quasi-mathematical approach to syntax. Focussing on a recent variant of this theory by Ray Jackendoff, I argue that this program is compatible with Whorf's Linguistic Relativity Hypothesis. However, the program faces insuperable difficulties, or so I will argue.

Finally, a third program for the study of natural language, the "truth-conditional program", is examined. I claim that in spite of a number of difficult problems faced by this program, it is our most promising avenue. A major characteristic of the truth-conditional program is that it generates an account of natural language that only makes sense in the context of a larger theory of social action; i.e. a sociological theory of the type that Max Weber advocated. If the Linguistic Relativity Hypothesis is to be viable as an empirical hypothesis, it must be so within the context of such a theory.

Chapter Three - Thought

The dependent variable in Whorf's hypothesis is thought. In this chapter I argue that thought, or more specifically, the content of thought, should be represented by what philosophers call the "propositional attitudes"; e.g., the belief that p, where p is a proposition (i.e., something that can be said in a sentence). A major virtue of this approach is that is consistent with the Weberian theoretical context that we have already endorsed in our truth-conditional approach to natural language. Alternative claims that the content of thought is represented non-propositionally are discussed and rejected.

The propositional-attitude account of the content of thought is the only account of the content of thought. But is it "scientific"? In opposition to theorists like Jerry Fodor, I argue that it is not; there are striking divergences in how we employ the propositional attitude idiom, as compared to how we use the language of physics or biology. On the other hand, I argue against theorists like Stephen Stitch, who argue that the propositional-attitude idiom should therefore be expunged from our conceptual scheme. Instead, I come out in favor of the view that the study of language and thought belong to the *Geisteswissenschaften*, which have a different set of epistemological norms than the *Naturwissenschaften*.

Chapter Four - Relativity

If the arguments of the previous chapters are correct, then to describe the language spoken in a community is to identify the truth-conditions of the utterances that members of that community might possibly make, and to describe the contents of the thoughts of the members of the community is to identify their propositional attitudes, i.e., their beliefs, desires, etc.

In this chapter I consider what method we might use in order to identify these truth-conditions and propositional attitudes. I argue that neither linguistic nor psychological facts can be discovered independently of one another. Both types of facts must be generated simultaneously by means of a method of interpretation. Interpreting behavior is like solving a set of equations with two unknowns.

I then review an argument by Donald Davidson that this joint interpretive method has the consequence that no community can have beliefs that are radically different from the beliefs of any other community. Davidson's argument entails that the Linguistic

Relativity Hypothesis is false, not empirically false, but false in a "conceptual" sense.

However, I go on to argue that Davidson overdoes the case for conceptual commonality. I consider a number of criticisms of Davidson's argument, each of which, if correct, will license a greater latitude of variation in cognitive attributions. In particular, I will show that it is possible to attribute cognitive variation to people by showing either that their referential systems vary, or that their standards of rationality vary (within limits). However, I will argue that both "relativity of reference" and "relativity of reasoning" result from public, historical events that are consciously recognized by members of the community. This is a very different picture than Whorf paints. Whorf claims that language results in cognitive variation through some kind of unconscious influence on language users. I argue that we can make no sense of these supposed unconscious influences.

In the end, then, we must reject Whorf's Linguistic Relativity Hypothesis.

CHAPTER TWO - NATURAL LANGUAGE

2.1 INTRODUCTION

Our understanding of the Linguistic Relativity Hypothesis is only as good as our understanding of language. The purpose of this chapter is to examine the concept of natural language and to determine whether a defensible concept of natural language can be consistent with Whorf's thesis.

Two of Whorf's key assumptions about natural language are the following:

1. Each language has a unique set of grammatical categories.

One can either be a realist or an instrumentalist about grammar. An instrumentalist holds that grammars are merely theoretical tools for describing the real facts. The "real facts" of linguistics, according to the instrumentalist, are described at a physical or perhaps physiological level, and they consist of descriptions of the sequences of sounds that the members of a speech community tend to produce. An instrumentalist holds that any grammatical description of a language that have the same consequences in terms of the "real facts" are equivalent.

The grammatical realist, on the other hand, thinks that there are facts about grammars; that is, grammars are real and can have causal consequences. The purpose of a grammatical description, on this view, is to directly characterize this grammatical reality. Therefore the realist is inclined to say that there is one particular grammar that best describes a natural language. Whorf was a grammatical realist in this sense.

2. The grammatical categories of a natural language have semantic correlates.

Whorf thought that the grammatical categories of a natural language have semantic correlates. A crude example of what Whorf was getting at is the homily that a noun (in English, at least) characterizes "a person, place or thing". A grammatical category, e.g., noun, is thus taken to have a "meaning" of sorts. The entire set of semantic correlates are what Whorf means by the conceptual structure of the language.

Strategy of the Chapter

This chapter will be directed at examining the concept of natural language in order to determine whether Whorf's assumptions are viable. My strategy will be to first examine the structuralist linguistics of Whorf's time. I will conclude that this approach to linguistics cannot generate the semantic descriptions that Whorf required for the Linguistic Relativity Hypothesis. I also claim that Whorf did not provide any clear account of how the structuralist program can be modified or enhanced to meet his requirements.

Attention will then be turned to the semantic theories offered by generative grammarians in the sixties. The generative approach to semantics seems to conform quite well to Whorf's rather sketchy writings on semantics. I will make the case that this is the semantics that Whorf wanted, but never had. However, this program will be criticized on a number of grounds, and ultimately rejected.

Finally the truth-conditional approach to semantics will be described and defended. This approach does give us a viable way to do semantics, or so I will argue. If Whorf's Linguistic Relativity Hypothesis is a viable hypothesis, it must be formulated within the context of a truth-conditional approach to natural language semantics.

Terminology

I have been freely using terminology such as 'grammar', 'syntax' and 'semantics'. These terms are used in various ways by different linguists and philosophers; consequently, I will point out some of the various usages, and indicate which usages I will be employing.

A good place to start is with the set of three terms introduced by Charles Morris[48]. Morris said that the study of language could be divided into three areas called syntax, semantics and pragmatics. Syntax, according to Morris, is the study of how sentences are formally structured from a stock of sub-sentential elements. Semantics is the study of the relation between language and the objects to which its expressions are applicable (roughly, the study of "meaning"). Finally, pragmatics is the study of the relation between language and its interpreters; in other words, it is the study of how language is used by individuals and communities.

It is very common to further differentiate additional areas of study within what Morris called syntax. In a typical linguistics textbook, written by the structuralist linguist Norman Stageberg[49], there are major sections on phonology, morphology, and syntax (where 'syntax' in Stageberg's sense is obviously more narrow in scope than Morris'). These sections correspond to Stageberg's view (shared by many linguists) that there are three levels of formal structure (or 'syntax', in Morris' sense) in a natural language.

The idea of a phonological level is based on the assumption that each natural language defines a finite set of sound-classes, commonly called 'phonemes', such that each phoneme is recognized as a distinct sound by speakers of that language. Each phoneme can be thought of as a set consisting of the physical sounds that can count as a token of that phoneme. The field of phonology, then, is the study of the phonemes of natural languages, specifically, what the phonemes are for a particular natural language, and how they interact.

---

[48] C. Morris, "Foundations of the Theory of Signs", in International Encyclopedia of Unified Science, ed. by O. Neurath, R. Carnap and C. Morris (University of Chicago Press, 1938), 84-85.

[49] N. C. Stageberg, An Introductory English Grammar, (Holt, Rinehart and Winston, 1965).

The morphological level deals with, roughly, the internal structure of words. More precisely, it is study of the formal, internal structure of morphemes, which are defined as follows:

> A morpheme is a short segment of language that meets three criteria:
> 1. It is a word, or a part of a word that has meaning.
> 2. It cannot be divided into smaller meaningful parts without violation of its meaning or without meaningless remainders.
> 3. It recurs in different verbal environments with a relatively stable meaning.[50]

Morphology, then, is the study of how morphemes are constructed out of the phonemic "alphabet" of the language, of how morphemes can be formed out of other morphemes, and so on.

Finally, the level of syntax, according to structuralist linguists like Stageberg, deals with how morphemes are structured into sentences.

In the late fifties Noam Chomsky launched a new era in linguistics, introducing a method of theorizing in linguistics that we now call generative grammar. With this new era came a new way of using terms.[51] 'Phonology' retained the same sense that it had with the structuralist linguists. However, the generative grammarians used 'syntax' to cover everything that the structuralists had included under morphology and syntax. Finally, 'grammar' was used to describe the combined theory of phonology, syntax, and semantics.

---

[50] Ibid., 85.

[51] This usage is explained in N. Chomsky, Aspects of the Theory of Syntax, (The M.I.T. Press, 1965), 15-16.

These terminological variations are summarized in the following table:

Table 2.1.1
Comparison of Terminology Used by Various Theorists

	Charles Morris	Structuralist Linguists	Generative Grammarians	
Sounds	Syntax	Phonology	Phonology	Grammar
Words		Morphology	Syntax	
Sentences		Syntax		
Meaning	Semantics		Semantics	
Use	Pragmatics			

In what follows, I will generally use these terms in the following manner:

Table 2.1.2
Terminology Used in the Present Work

	Terminology		
Sounds	Phonology		Grammar
Words	Morphology	Syntax	
Sentences	Sentence-syntax		
Meaning	Semantics		
Use	Pragmatics		

On occasions where I revert to alternate usage I will make this clear in the context.

2.2 BLOOMFIELD'S LINGUISTIC PROGRAM

American linguistics in the period from 1925 to 1956 was dominated by an approach variously called "structuralist" or "descriptive" linguistics.[52] Whorf was a participant in this tradition, although as mentioned in Chapter One, he was not entirely satisfied with the structuralist position on semantic issues. This section will examine the structuralist program in more detail. As indicated in Chapter One, in order to formulate his Linguistic Relativity Hypothesis, Whorf required a concept of natural language such that each natural language has its own semantic structure. The purpose of this section, then, is to determine whether the structuralist program can deliver the type of analysis of natural language that Whorf requires.

---

[52] Many of the key papers in this tradition are collected in M. Joos, ed., Readings in Linguistics I, (University of Chicago Press, 1957).

Perhaps no other individual played a greater role in characterizing the objectives and assumptions of the structuralist program than Leonard Bloomfield (1887-1949).[53] I will focus on Bloomfield's writings, but I will also address other contributors to the structuralist tradition where appropriate.

Bloomfield on Where to Begin

> The most difficult step in the study of language is the first step. Again and again, scholarship has approached the study of language without actually entering upon it. Linguistic science arose from relatively practical preoccupations, such as the use of writing, the study of literature and especially of older records, and the prescription of elegant speech, but people can spend any amount of time on these things without actually entering upon linguistic study.[54]

If the analysis of writing systems, of literature, or the prescription of linguistic norms are not taking us to the heart of the matter, what is the heart of the matter? Bloomfield is unequivocal: to understand what natural language is we must begin with the use of language, or what Morris called "pragmatics". That is, we must begin with an examination of how spoken utterances are used by human beings in their daily transactions. "Language" must be characterized in terms of the more fundamental, more observable phenomenon of speech-utterances.[55]

---

[53] The best sources for an account of Bloomfield's theoretical position are: L. Bloomfield, "A Set of Postulates for the Science of Language", Language, 2, 1926, 153-164 (reprinted in Joos, ibid.); and the very influential: L. Bloomfield, Language, (1933, reprint University of Chicago Press, 1984).

[54] L. Bloomfield, Language, (1933, reprint University of Chicago Press, 1984), 21.

Bloomfield asks us to compare two situations. (A) Jill, who is hungry, sees an apple. She then obtains and eats the apple. (B) Jill, who again is hungry, is with Jack. Jill produces some sounds. Jack obtains the apple and gives it to Jill, who eats it.

Bloomfield, who was very influenced by the behavioristic terminology and thinking of his day, uses the following diagrams to analyze the two situations.

$$\text{(A)}\quad S \longrightarrow R$$

$$\text{(B)}\quad S \longrightarrow r \;.........\; s \longrightarrow R$$

Situation (A), where no speech is involved, is characterized by Jill being visually stimulated by light reflecting off an apple (S), and her subsequent response (R), which consisted of obtaining and eating the apple. The arrow connecting the S and the R represents activities in Jill's nervous system.

Situation (B) is characterized by Jill's being visually stimulated by the apple (S), but this time it leads to a speech response on Jill's part (r). The arrow between S and r represents activities in Jill's nervous system. The dotted line represents the physical consequences of Jill's speech, that is, movement of the air around Jill. As a result of this movement of air, Jack receives an auditory stimulus (s), and as a result, Jack responds by obtaining the apple and giving it to Jill (R). The arrow connecting the s and the R represents activities in Jack's nervous system.

---

[55] Ibid., 22.

According to Bloomfield, in order to have a scientific account of language we must forget all our presumptions about language, our interest in writing systems, our interest in literature and other sophisticated uses of language, and our tendency to think of language in prescriptive or normative terms. Instead, we should adopt an objective and materialistic[56] attitude to situations like (B). The study of language is nothing more than the study of situations like (B). In fact, the linguist should focus his study on just one component of these situations.

> In the division of scientific labor, the linguist deals only with the speech-signal (r........s); he is not competent to deal with problems of physiology or psychology. The findings of the linguist, who studies the speech-signal, will be all the more valuable for

---

[56] "The materialistic (or better mechanistic) theory supposes that the variability of human conduct, including speech, is due only to the fact that the human body is a very complex system. Human actions, according to the materialistic view, are part of cause-and-effect sequences exactly like those which we observe, say in the study of physics or chemistry. However, the human body is so complex a structure that even a relatively simple change, such as say, the impingement on the retina of light-waves from a red apple, may set off some very complicated chain of consequences, and a very slight difference in the state of the body may result in a great difference in its response to the light-waves. We could foretell a person's actions (for instance, whether a certain stimulus will lead him to speak, and, if so, the exact words he will utter) only if we knew the exact structure of his body at the moment, or, what comes to the same thing, if we knew the exact makeup of his organism at some early stage - say at birth or before - and then had a record of every change in that organism, including every stimulus that had ever affected the organism." Ibid., 33.

> the psychologist if they are not distorted by any prepossessions about psychology.[57]

The linguist's job, then, is to focus on the speech-signal (r........s), whereas the psychologist of language deals with the other components of speech behavior, namely, (S ————————> r) and (s ————————> R). So we see that although the linguist must start with speech episodes in order to get a clear understanding of what his subject matter is, the linguist need not, and should not concern himself with all aspects of speech-episodes. This doctrine can be called "the autonomy of linguistics", and was extremely influential in American linguistics throughout the entire structuralist era and beyond.

Bloomfield tells us that as linguists, we are to focus our attention on the speech signal, which includes the formation of sounds by a speaker, the transmission of those sounds through the air, and the auditory reception of those sounds by a hearer. Whatever language is, it is to be discovered by focusing on these phenomena, and these phenomena alone. Now Charles Morris has told us that there are three divisions in the study of language: syntax, semantics and pragmatics. Let us now consider what Bloomfield's prescription implies about each of Morris' divisions.

First, pragmatics. Pragmatics, or the study of the use of language, would obviously require consideration of all aspects of speech episodes like (B). Since on Bloomfield's view the linguist can only contribute a portion of a complete analysis of a speech episode, the bulk of the study of pragmatics falls to the psychologist (or perhaps Bloomfield would approve of a behavioristically oriented sociologist) rather than the linguist.

---

[57] Ibid., 32.

What about semantics, or the study of linguistic meaning? The following well-known passage from Bloomfield is worth quoting at length:

> The study of speech sounds without regard to meanings is an abstraction: in actual use, speech-sounds are uttered as signals. We have defined the *meaning* of a linguistic form as the situation in which the speaker utters it and the response which it calls forth in the hearer. The speaker's situation and the hearer's response are closely coordinated, thanks to the circumstance that every one of us learn to act indifferently as a speaker or as a hearer. In the causal sequence
>
> speaker's situation --> speech --> hearer's response,
>
> the speaker's situation, as the earlier term, will usually present a simpler aspect than the hearer's response; therefore we usually discuss and define meanings in terms of a speaker's stimulus.
>
> The situations which prompt people to utter speech, include every object and happening in their universe. In order to give a scientifically accurate definition of meaning for every form ['form' is Bloomfield's term for words, phrases, etc.] of a language, we should have to have a scientifically accurate knowledge of everything in the speaker's world. The actual extent of human knowledge is very small, compared to this. We can define the meaning of a speech-form accurately when this meaning has to do with some matter of which we possess scientific knowledge. We can define the names of minerals, for example, in terms of chemistry and mineralogy, as when we say that the ordinary meaning of the English word *salt* is 'sodium chloride (NaCl),' and we can define the names of plants or animals by means of the technical terms of botany or zoology., but we have no precise way of defining words like *love* or *hate*, which concern situations that have not been accurately classified - and these latter are the great majority.[58]

---

[58] *Ibid.*, 139.

Bloomfield's stance on semantics involves two key points.

(1) Semantic phenomena involve more than the speech signal (r........s). The meaning of a speech signal, according to Bloomfield, has to be characterized in terms of the entire speech episode. Consequently, this means that for Bloomfield there is no real distinction between pragmatics and semantics. Furthermore, it means that semantics is not really the province of the linguist, qua linguist. Semantics involves psychological factors, as well as whatever sciences are required to describe the environmental factors responsible for the speaker's stimulus.

(2) Bloomfield states that semantics has to be put on hold until significant advances are made in other sciences. This point is, of course, open to debate. In fact, I will challenge this assumption later on. However, Bloomfield's stance was very influential, and the majority of American structuralist linguists accepted Bloomfield's stance, and felt that semantic issues could be put neatly to the side for the time being.

That leaves syntax, which in Morris's sense includes phonology, morphology, and sentence-syntax. It is syntax, in exactly this sense, that Bloomfield considered the proper domain of the linguist. Bloomfield believed that the linguist can and should study syntactic phenomena independently of semantic/pragmatic phenomena.

In summary, Bloomfield believed that in order to really grasp what natural language is, we have to see it as an abstraction from actual speech events. The concept of language applies to the speech-signal as a component of speech events, and speech events are events classified in pragmatic/semantic terms. However, once the linguist is properly focused on speech events, Bloomfield says that he should ignore all semantic and pragmatic considerations and focus solely on the speech-signal. In other words, Bloomfield held that the semantic/pragmatic context is required to characterize language in general, but that it should be ignored when studying specific languages in detail.

Bloomfield on The Description of a Natural Language

Bloomfield's classic article "A Set of Postulates for the Science of Language"[59] identifies the "picture" of language that he advocated, and which was adopted by most of the descriptive linguists that followed him. I will now present the ideas of Bloomfield's "Postulates"; however, I will use more contemporary terms when appropriate.

Bloomfield imagines a group of people who can verbally communicate with one another. This he calls a speech-community. They communicate through speech-signals or utterances. The totality of possible utterances that can be made by the individuals within a speech-community is a language.

---

[59] L. Bloomfield, "A Set of Postulates for the Science of Language", Language, 2, 1926, 153-164. Reprinted M. Joos, ed., Readings in Linguistics I, (University of Chicago Press, 1957). In his editorial comments, Martin Joos states that "The Postulates have been called the Charter of contemporary descriptive linguistics. Nothing in them has been truly superceded, though much of Bloomfield's other work has been, a fate which he regarded as perfectly natural and indeed inevitable." p. 31. Joos' commentary was written in the late fifties; the golden age of structuralist linguistics, just as the Chomskyan onslaught was about to begin.

An utterance consists of one or more sentences. A sentence is a unit of speech which "is not part of a larger construction". It follows from this, and the last sentence of the previous paragraph, that you will specify a language if you can specify all the sentences.[60]

Utterances have meanings. These meanings are to be explained in terms of stimulus-response psychology. As noted, Bloomfield was very sceptical that a complete semantic description could be provided in the short term; however he held that we do need a few semantic principles in order to make progress on syntactic descriptions. In particular, we need the notion of a significant vocalization, for this is what constitutes a sentence-utterance and separates it from vocal nonsense. Sentence-syntax is the purely formal study of how sentences are put together from sub-sentential elements, but we can make no progress on this task unless we know what the sentences are, and only semantic judgments can tell us that. Furthermore, we need the notion of two utterances being the same in meaning. (It will be shown why this is necessary later on when the discovery procedures are examined.) So Bloomfield holds that in order to specify the syntax of a natural language two general semantic notions are required: the idea of a significant vocalization and the idea of a synonymy relation defined over the class of sentence-utterances.

We have seen that Bloomfield attaches special significance to the sentence, as it is a minimal significant utterance.

---

[60] Ibid., 26.

However, Bloomfield holds that sentences have parts, and these parts have *meanings* as well. These, along with sentences, are called *forms*. The non-sentential forms are called *morphemes*. The concept of a morpheme is fairly close to the everyday concept of a word, which is to say that it is a fundamentally semantic notion. That is, the morphemes of a language are the smallest units that have an independent meaning. They do not have meaning in the sense that they can be used to make utterances. As we have seen, Bloomfield uses the term 'sentence' to describe the smallest units that can be used to make utterances. Rather, morphemes have meanings in the sense that each morpheme makes a similar semantic contribution to each sentence in which it appears. For Bloomfield, then, there is a fundamental distinction between word-meaning and sentence-meaning. Bloomfield held that in order to specify the syntax of a natural langauge it is necessary to first identify all the morphemes and a number of sentences of the language. In order to do this the linguist must therefore make a number of semantically based judgments. However, as we have seen, Bloomfield was doubtful about our ability to construct detailed semantic theories at this time. His guiding principle was to avoid semantics and semantically based judgements as much as possible; however, he realized that some of these judgments are necessary in order to generate the data the syntactician needs to do his work.

Bloomfield also endorsed the so-called phonemic principle. The idea is that for each language there is a finite and relatively small set of distinctive sounds that are implicitly recognized as distinct units by the speakers of that language. These are called the phonemes of the language. All utterances consist solely of strings of phonemes.[61]

One might think that phonemes can be defined as equivalence classes of sounds, physically described. However, Bloomfield, along with almost all of his contemporaries, held that there is another level of analysis between the physical or acoustic description and the phonemic description. This is the phonetic level. The idea is that all possible vocal sounds can be classified into a finite number of phones. The class of phones is larger than the class of phonemes, but still small enough that it can be mastered as a practical tool for linguists. The phonemes of a particular language can then be defined as equivalence classes of phones. What are phones? They should be understood as a practical tool for the field linguist. Think of them as the smallest possible set of distinctive sound groupings that would be sufficient for producing a phonemic analysis of any possible human language.

It might be objected that we cannot know in advance what natural languages are possible and that therefore phonetic analysis lacks a solid empirical foundation. This is perhaps correct, but it does not negate the practical utility of phonetics. Think of phonetic analysis in instrumental terms. It would be possible for a linguist to adhere to the phonemic principle without the fiction of phones, for he could think of phonemes as equivalence classes of raw sounds, acoustically described. However, it would be hugely inconvenient to do so in practise, since he would have to go to all the bother of characterizing the acoustic boundaries. How much more convenient to just do this job once, marking the possible boundaries that can occur in any possible language. This is what phonetic analysis does, and it can be easily appreciated that it saves a huge amount of effort. Note, however, that any particular scheme of phonetic analysis is based on an empirical hypothesis, namely, that human beings have a tendency, no matter what their mother tongue, to organize sounds into the "phones" that are postulated by that particular phonetic scheme.

---

[61] Ibid., 27. For more on the phonemic principle see M. Swadesh, "The Phonemic Principle", Language, 1934, 10, 117-129, reprinted in Joos, ibid.

Linguists have worked out a set of phones that they think are appropriate for the description of all possible utterances in all possible languages. This is called the International Phonetic Alphabet.[62] The idea is that a field linguist can go into a speech-community who use a language that he does not understand, and he can use the International Phonetic Alphabet to collect data. That is, he can phonetically transcribe the utterances that he hears. This phonetic transcription can be used as the base data on which later linguistic analysis can be performed. The first step will be to come up with a phonemic transcription appropriate for that particular language. That is, the linguist will attempt to determine which phones are indistinguishable for the members of the speech community. Indistinguishable phones will be collected together into equivalence classes called phonemes. The original phonetically transcribed corpus of data can now be phonemically transcribed.

---

[62] The alphabet is described in: The International Phonetic Association, The Principles of the International Phonetic Association, published by the Association as a pamphlet, 1949. Although this pamphlet was published much later than Bloomfield's seminal works, the International Phonetic Alphabet was produced earlier. Bloomfield made use of it in his 1933 Language, and he mentioned several other competing phonetic alphabets. He remarked, "In principle, one phonetic alphabet is about as good as another, since all we need is a few dozen symbols, enough to supply one for each phoneme of whatever language we are describing." p. 87.

This newly transcribed corpus can now be used for higher levels of structural analysis, namely the identification of morphemes and of constructions. The identification of morphemes is called morphology, and the identification of constructions is called sentence-syntax.

The previous paragraph suggested that there are "levels" of syntactic analysis, and that is exactly what Bloomfield believed. Specifically, he held that each natural language has to be described at three distinct syntactic levels:

1. The phonological level - This level is a specification of the phonemes of the language, where phonemes are equivalence classes of phones. Phonemes represent the classes of sounds that are distinguished by members of that speech community.

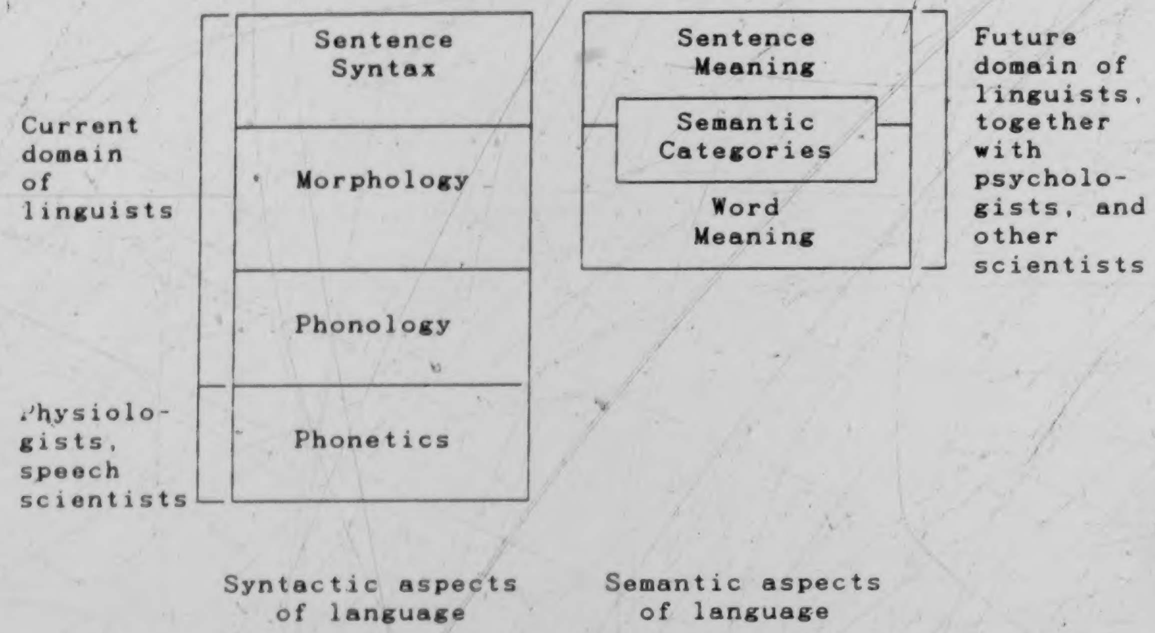
2. The morphological level - This level is a specification of the morphemes (i.e. words and formatives) of the language. Furthermore, the morphological description of a language must include the set of morpheme construction rules that specify how new morphemes can be formed from phones. These rules implicitly define the set of possible morphemes for that language.

3. The sentence-syntax level - Not any stringing together of morphemes will constitute a sentence of the language. The idea of the sentence-syntax level is to identify the acceptable "constructions" of the language.

Bloomfield held that it is the job of the contemporary linguist to specify all three levels of syntax for particular natural languages. The linguist of the future will also have to construct a semantic theory, but this should be avoided for the foreseeable future. Today's linguist must rely on semantically based judgements in order to determine what strings of sounds constitute sentences and morphemes (and ultimately, as will be shown in the discussion of "discovery procedures" to follow, this reliance means that phonemes are also characterized in terms of semantic judgements); however, the reliance on semantic judgments should be kept to a bare minimum.

Bloomfield's overall "paradigm" for the study of language is diagrammed below.

Figure 2.2.1
Bloomfield's Paradigm for the Study of Natural Language


Current domain of linguists
Physiologists, speech scientists
Sentence Syntax
Morphology
Phonology
Phonetics
Sentence Meaning
Semantic Categories
Word Meaning
Future domain of linguists, together with psychologists, and other scientists
Syntactic aspects of language
Semantic aspects of language

Note that Bloomfield believed that the semantic description of a language should include an account of the language's "semantic categories" as well as its sentence meanings and word meanings. Bloomfield proposed the following definition:

> The functional meaning and the class meanings of a language are the categories of the language.[63]

A "functional meaning" is a meaning associated with a "position" in a syntactic construction. For example, the sentence 'Richard saw John' is a syntactic construction that has three positions: actor ('Richard'), predicate ('saw'), and object ('John'). Bloomfield pointed out that a particular morpheme typically can occupy more than one position; for example, 'Richard' and 'John' are interchangeable in position. Bloomfield defined "form-class" as any class of morphemes such that each member of the class can occupy the same positions. The notion of form-classes is approximately equivalent to the common notion of "parts of speech", that is, the classification of words into nouns, verbs, adverbs, adjectives, etc. Each form-class has a meaning associated with it, and these meanings are called "class meanings". A complete semantic theory must include an account of the categories, that is, the position meanings and the class meanings. Note that Whorf's Linguistic Relativity Hypothesis is basically the claim that Bloomfieldian categories determine the way that language users think. Presumably Bloomfield would consider such a hypothesis to be premature, since it is not possible, in his opinion, to formulate a detailed semantic theory at this time.

---

[63] Bloomfield, "A Set of Postulates for the Science of Language", Language, 2, 1926, 153-164. Reprinted M. Joos, ed., Readings in Linguistics I, (University of Chicago Press; 1957), 29.

We will return to Whorf's relation to the Bloomfieldian paradigm shortly. First, there will be brief examination of Bloomfield's methodological prescriptions.

The Discovery Procedures

Bloomfield was one of many American scholars of the early twentieth century who reacted strongly against introspective techniques in psychology, linguistics and other sciences of man. Greatly influenced by the operationalist philosophy of science, Bloomfield felt that every term introduced into any science, including the science of linguistics, had to be supported by rigorous empirical "operations".

What this meant is that Bloomfield had to support his "picture" of language by a set of empirical operations that would fill out the picture for a particular speech-community. These operations were called "discovery procedures". The idea of the discovery procedures is as follows: The linguist starts with a physical description of the vocalizations within a speech community. He then applies certain objective, almost mechanical, techniques. As a result of applying these techniques (discovery procedures), he ends up with a three-layered syntactic description of the phonology, morphology and sentence-syntax of the language. The resulting syntactic description is entirely objective, since the linguist was merely following mechanical procedures that can, in principle, be followed by anyone else.

Before the linguist can begin applying the discovery procedures he must have a body of data. This initial body of data has come to be called the "corpus". A corpus of linguistic data is a set of vocalizations, phonetically described. However, not any set of vocalizations will do. What the linguist needs in order to start applying the discovery procedures is those vocalizations that are utterances of single sentences. In order to judge whether a vocalization is an utterance of a single sentence (as opposed to the an utterance of many sentences, or an utterance of a part of a sentence, or merely the making of a non-linguistic noise) the linguist must employ a pragmatic/semantic notion of "significance". That is, a "significant" vocalization is one that can be used to perform a complete speech act. Utterances of sentences, then, are the shortest vocalizations that are significant. Furthermore, the corpus must be characterized by another semantic notion, that of synonymy. More precisely, the corpus, which is a set consisting of utterances of sentences, must also have a synonymy relation defined on it. Synonymy, or likeness of meaning, is an equivalence relation such that the relation will divide the corpus into a number of equivalence classes. Each equivalence class will represent a sentence, and each member of an equivalence class is an utterance of that sentence, perhaps with variations in pronunciation.

In summary, the "corpus" from which the structuralist linguist begins is a set of phonetically described utterances where each member of the set is an utterance of a sentence, and where the utterances are classified in terms of which sentences they are utterances of. In order to produce such a corpus it is necessary to rely on two semantic notions, significance and synonymy.[64]

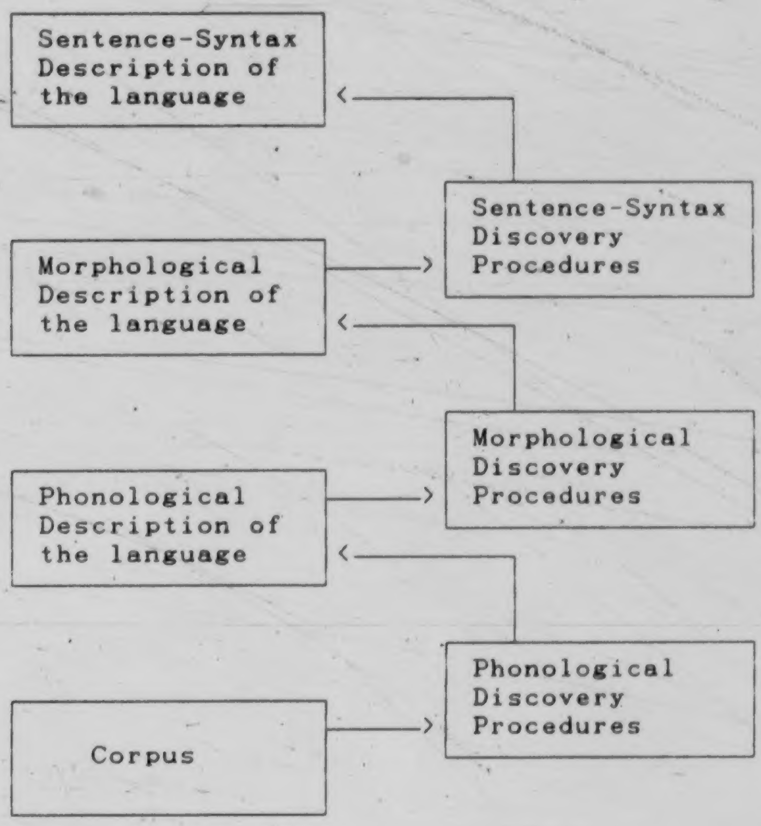
If such a corpus has been produced, presumably through the work of a field linguist, and assuming it is of sufficient size, then according to the Bloomfieldian doctrine all one has to do is apply the discovery procedures in order to produce a complete syntactic description of the language. No further semantic or pragmatic facts have to be introduced. Theoretically, the discovery procedures can be applied by the linguist working at his desk, far from the speech-community from which the corpus was derived (although Bloomfield, Harris and the others realized that this would not be a practical way of doing linguistics).

---

[64] The reliance on these two semantic notions is implicit in the work of Bloomfield and other important structuralist linguists. See especially Z. S. Harris, Structural Linguistics, (University of Chicago Press, 1951). W. V. O. Quine makes this reliance explicit in "The Problem of Meaning in Linguistics", in From a Logical Point of View, 2nd ed., (Harper and Row, 1961).

There are three distinct types of discovery procedures, corresponding to the three levels of syntactic analysis. As indicated in the following diagram, the discovery procedures are applied in a uni-directional manner; first the phonological description is generated, then the morphological, and finally the description of sentence-syntax is produced.

Figure 2.2.2
Discovery Procedures


Sentence-Syntax Description of the language
Sentence-Syntax Discovery Procedures
Morphological Description of the language
Morphological Discovery Procedures
Phonological Description of the language
Phonological Discovery Procedures
Corpus

There are a number of principles and terms that are common to all three types of discovery procedures. It will facilitate the discussion to review these common principles and terms before turning to the detailed procedures.

First of all, the members of the corpus are considered to be strings of elements. Each level of syntactic description has its own unique elements. At the level of the unanalyzed corpus the elements are phones, at the phonological level the elements are phonemes, at the morphological level they are morphemes, and at the sentence-syntax level they are morpheme classes. Furthermore, the elements at level n are either sets or strings of elements defined at level n-1. Thus:

sentence-types are strings of morpheme classes,
sentences are strings of morphemes,
morpheme-classes are sets of morphemes,
morphemes are strings of phonemes, and
phonemes are sets of phones.[65]

The purpose of the discovery procedures for level n is to take the corpus as input, where the corpus is represented as strings of elements of level n-1, and to generate as output all the elements of level n, as well as a new representation of the corpus as strings of elements of level n. In order to do this the discovery procedures must attend to the formal distribution pattern of elements in the corpus. The distribution of an element is simply the set of environments in which it can occur.[66]

---

[65] The claim that the elements of one level of analysis are sets or strings of elements at a lower level is one of the main tenets of Bloomfield's program. It is called the "taxonomic condition" in J. A. Fodor, T. G. Bever, and M. F. Garrett, The Psychology of Language, (McGraw-Hill, 1974), 30. These authors point out that "its effect is to order the descriptive levels posited by a taxonomic grammar into a hierarchy of classes."

[66] "The ENVIRONMENT or position of an element consists of the neighborhood, within an utterance, of elements which have been set up on the basis of the same fundamental procedures which were used in setting up the element in question.... The DISTRIBUTION of an element is the total of all environments in which it occurs, i.e., the sum of all the (different) positions (or occurrences) of an element relative to the occurrences of other elements." These definitions are from Z. S. Harris, *Structural Linguistics*, (University of Chicago Press, 1951), 15-16.

These concepts are best explained by an example. Suppose at some level of analysis there are exactly two elements: a and b. Now consider the set E = {a, b, _}, where _ is a place-holder symbol. Now we can form a set of "environments" which contains all the strings that can be obtained by concatenating the members of E, providing that there is exactly one place-holder symbol in the string. Thus, the environments that can be formed from E will include:

Figure 2.2.3
"Environments" based on the Vocabulary 'a', 'b'

_a	_aa	_aaa	_aaaa	_aaaaa	... etc.
_b	_ab	_aab	.	.	
a_	_ba	_aba	.	.	
b_	_bb	_abb	.	.	
	a_a	_baa			
	a_b	_bab			
	b_a	_bba			
	b_b	_bbb			
	aa_	a_aa			
	ab_	a_ab			
	ba_	a_ba			
	bb_	a_bb			
		b_aa			
		.	.	.	.
		.	.	.	.
		etc.	etc.	etc.	etc.

Now suppose our corpus for this language consists of five members: (1) ab, (2) bb, (3) aba, (4) baa, and (5) bab. The distributional analysis of the elements in this corpus can be represented by means of a matrix as follows:

Figure 2.2.4
Matrix for Representing a Corpus of Data

		Elements	
		a	b
Environments	_a		
	_b	(1)	(2)
	a_		(1)
	b_		(2)
	_aa		(4)
	_ab		(5)
	_ba	(3)	
	_bb		
	a_a		(3)
	a_b		
	b_a	(4)	
	b_b	(5)	
	aa_		
	ab_	(3)	(4)
	ba_		(5)
	bb_		
	.		
	.		
	etc.		

There is a column in the matrix for each element, and there is a row for every environment that can be formed from those elements. The matrix cells are filled in as follows: if, by placing the element corresponding to that cell's column into the place-holder position of the environment corresponding to that cell's row, a corpus member is generated, then place an index number that has been arbitrarily asssigned to that corpus member into the cell; otherwise leave the cell blank. Clearly, all the facts about the formal distribution of elements in the corpus members will be fairly obviously represented by such a matrix.[67]

Structuralist linguists devised a number of terms to describe various aspects of distribution patterns. Two elements are said to be in complementary distribution if they never appear in the same environments. In the example given above, a and b are not in complementary distribution, because both can occur in the environments _b and ab_.[68] However, it is fairly common to observe pairs of elements that are in complementary distribution. At the phonological level, for example, the phonemes /es/ and /z/ are in complementary distribution in English. 'Foxes' and 'boxes' end with the /z/ phoneme whereas 'books' and 'stamps' end with the /es/ phoneme. In environments where one phoneme appears, the other never does. Similarly at the morphological level many morphemes (words) are in complementary distribution with each other.

---

[67] In set theoretic terms, the matrix is a representation of EL X EN -> C, where EL is the set of elements, EN is the set of environments, and C is the corpus. That is, the matrix is a function from the Cartesian product of the set of elements and the set of environments into the corpus. However, the reader should be aware that this formalization of distribution analysis in terms of the construction of such a matrix is for instructive purposes only. Such a matrix would be too large and unwieldy to be practical. Instead, what structural linguists typically presented in their papers were tables of data that can easily be construed as the most significant portions of such a matrix. For example, see Harris' list of consonants of Swahili in Z. S. Harris, Structural Linguistics, (University of Chicago Press, 1951), 99-109.

[68] In terms of the matrix approach, two elements will be in complementary distribution if and only if for each row of the matrix, at most one of the columns corresponding to the elements has an entry in that cell.

Another important notion is that of a minimal pair. A pair of elements form a minimal pair if by substituting one for another in any environment the resulting utterance is not synonymous with the first. Obviously, the notion of a minimal pair is not a purely distributional (or formal) notion; it is a partially semantic notion. Recall that when the corpus was first characterized, it was pointed out that a synonymy relation was defined on the corpus. Suppose that in our example corpus, (1) and (2) are synonymous, and (3) and (4) are synonymous. Are a and b a minimal pair? No, they are not. We can easily spot the environments where substitutions are possible. They are the environments whose matrix rows have entries in more than one column. Thus, a and b can both appear in two environments: _b and ab_. But note that if we substitute either element into _b we get either (1) or (2), which have been judged to be synonymous. And if we substitute either element into ab_ we get either (3) or (4), which also have been judged to be synonymous. Consequently, a and b are not minimal pairs. Whenever two elements do not form minimal pairs, they are said to be in free variation.[69]

---

[69] A typical statement of these key concepts is the following: "If different phones have none of their environments in common, they are IN COMPLEMENTARY DISTRIBUTION; if they have in common only a phonetically or phonemically definable set of environments, they are IN OVERLAPPING DISTRIBUTION; if one phone shares all the environments of another, they are IN FREE VARIATION." From B. Bloch, "Studies in Colloquial Japanese IV: Phonemics", Language, 1950, 26, 86-125. Reprinted M. Joos, ed., Readings in Linguistics I, (University of Chicago Press, 1957), 330.

Note that if our synonymy judgements had been different, then we would draw very different conclusions from the distributional data. Suppose that in our original corpus (1) and (2) were judged synonymous, (3) and (5) were judged synonymous, and (4) was judged non-synonymous with all the others. In that case, a and b would form a minimal pair, for in the environment ab_ they form the non-synonymous utterances (3) and (4) respectively. Since they do form a minimal pair, they are not in free variation.

We now have developed enough terminology to characterize the basic principles of the discovery procedures. Suppose we are analyzing level n. Then our input will be the corpus, where each member of the corpus is characterized as a string of elements of level n-1. Using the elements and environments of level-1, we form a matrix and enter the corpus members into the matrix. For each pair of level n-1 elements we examine the matrix and determine whether the pair is in complementary distribution. Then, for each pair of level n-1 elements we examine the matrix and the synonymy relation defined on the corpus and determine whether the pair is in free variation. Then we determine the elements of level n as follows:

> Each element of level n is a set of elements of level n-1 such that each member of the set is either in free variation with the others, or is in complementary distribution with the others and is somehow "similar" to the others.[70]

---

[70] In reducing the discovery procedures to the application of two rules (free variation and complementary distribution) I have followed the expository pattern of J. A. Fodor, T. G. Bever, and M. F. Garrett, The Psychology of Language, (McGraw-Hill, 1974), 28-50. A similar exposition is found in H. Maclay, "Overview -

If two level n-1 elements are in free variation, then they always belong to the same level n element. However, if two level n-1 elements are in complementary distribution, then they may belong to the same level n element, or they may not. The complementary distribution may simply be an accident. To rule out the accidental cases, structuralist linguistics generally look for some sort of "similarity" between the elements, where the similarity is measured in terms of some property defined at level n-1. An example may help. As pointed out above, /es/ and /z/ are in complementary distribution in English. Structuralist linguists use this fact to claim that there is a pluralization morpheme in English that has both of these single phoneme strings as members. To supplement the claim about complementary distribution, the structuralists point out the phonetic similarity between the two phonemes.

Phonological Discovery Procedures

Now let us briefly consider how these principles apply at each of the three levels of syntactic analysis. First, there are

---

Linguistics", in Semantics, edited by D. D. Steinberg and L. A. Jakobovits, (Cambridge University Press, 1971), 159-163, except that Maclay does not explicitly use the term 'free variation'. The writings of the original structuralist linguists who developed the discovery procedures are inevitably more complex and more subtle than the overview that is presented here. However, two classic articles that are fairly consistent with this presentation are M. Swadesh, "The Phonemic Principle", Language, 1934, 10, 117-129, and C. Hockett, "A System of Descriptive Phonology", Language, 1942, 18, 3-21. Both are reprinted in M. Joos, ed., Readings in Linguistics I, (University of Chicago Press, 1957). For an exposure to all the complexities of the discovery procedures, see Z. S. Harris, Structural Linguistics, (University of Chicago Press, 1951), especially the index entries "complementary" and "free variant". Harris' work is widely regarded as the ultimate formulation of this approach to linguistic methodology.

the phonological discovery procedures. The input is the corpus, phonetically transcribed (that is, transcribed as strings of phones), with a synonymy relation defined over the members of the corpus. A distribution analysis is performed; that is, a matrix is constructed with a column for each phone and a row for each phonetic environment. If by substituting a column's phone into a row's environment a corpus member is generated, then the index number of that corpus member is entered into the cell defined by that row and column. Then the phones are grouped into equivalence classes by using the free variation and complementary distribution tests. These equivalence classes of phones are the phonemes of the language. The corpus can now be phonemically transcribed.

Morphological Discovery Procedures

The next step is to apply the morphological discovery procedures. The input is the phonemic transcription of the corpus (of course, the synonymy relation will still be defined over the corpus members). Another distribution analysis is performed, except that this time the columns of the matrix will correspond to strings of phonemes, and the rows will correspond to phonemic environments. Then the linguist must identify the shortest possible phoneme strings such that when the string is placed in the range of phonemic environments a significant number of corpus members are produced. (In others words, identify the shortest strings of phonemes that have a lot of matrix entries in their columns.) Intuitively, these phoneme strings are "words" that reappear in different utterances. More precisely, these strings will be called the "morpheme alternants" of the language. However, this will result in 'knife' and 'knive_' being different morpheme alternants. Intuitively, we want to view them as the same word. So an additional procedure is applied: once the morpheme alternants have been identified, we look for patterns of complementary distribution between morpheme pairs. If a pair of morpheme alternants are in complementary distribution and are phonemically similar (as 'knife' and 'knive_' are), then they are assigned to the same morpheme, where a morpheme is to be understood as a set of morpheme alternants.

Once all the morphemes have been identified, a final step in morphological analysis is to examine the sequencing of phonemes in morpheme alternants in order to generalize some rules of morpheme construction. Typically, each natural language allows the formation of words with certain phoneme sequences but not others. It is the structural linguist's task to identify these restrictions by formulating a set of rules.[71]

[71] This presentation of morphological discovery procedures is based on Z. S. Harris, "Morpheme Alternants in Linguistic Analysis", Language, 1942, 18, 169-180. Reprinted M. Joos, ed., Readings in Linguistics I, (University of Chicago Press, 1957). However, in his paper Harris admits more semantic judgments into the corpus. Harris assumes that the corpus contains judgments regarding which sub-sentential phonemic sequences have "word-meaning", and he also assumes that a word-synonymy relation has been defined over the set of sub-sentential phonemic sequences. Thus, Harris ends up identifying four types of semantic judgments: sentence-significance, sentence-synonymy, word-significance, and word-synonymy. Presumably Harris introduced the last two judgments reluctantly, since along with other structuralist linguists he wanted to reduce his reliance on semantics to a minimum. However, I have argued that the latter two judgments are not required, since I have argued that "word-hood" can be identified purely on the basis of distributional data. However, this is ultimately an empirical claim, viz., that compared to non-morphemes, morphemes will generate far more

Once the morphological analysis is completed, the corpus can be transcribed into sequences of morphemes.

Sentence-syntax Discovery Procedures

The input to the sentence-syntax discovery procedures is the morphological transcription of the corpus. A distributional analysis is performed, with morphemes forming the columns of the distribution matrix, and morpheme-string environments forming the rows. Morphemes that are in free variation are assigned to the same morpheme-class. (Morpheme-classes are roughly equivalent to the "parts of speech", that is, nouns, verbs, etc.)

Then another type of distribution analysis is performed, where the columns of the matrix correspond to strings of morphemes, and the rows to morpheme-string environments. Strings of morphemes that are in free variation are assigned to the same phrase-type.

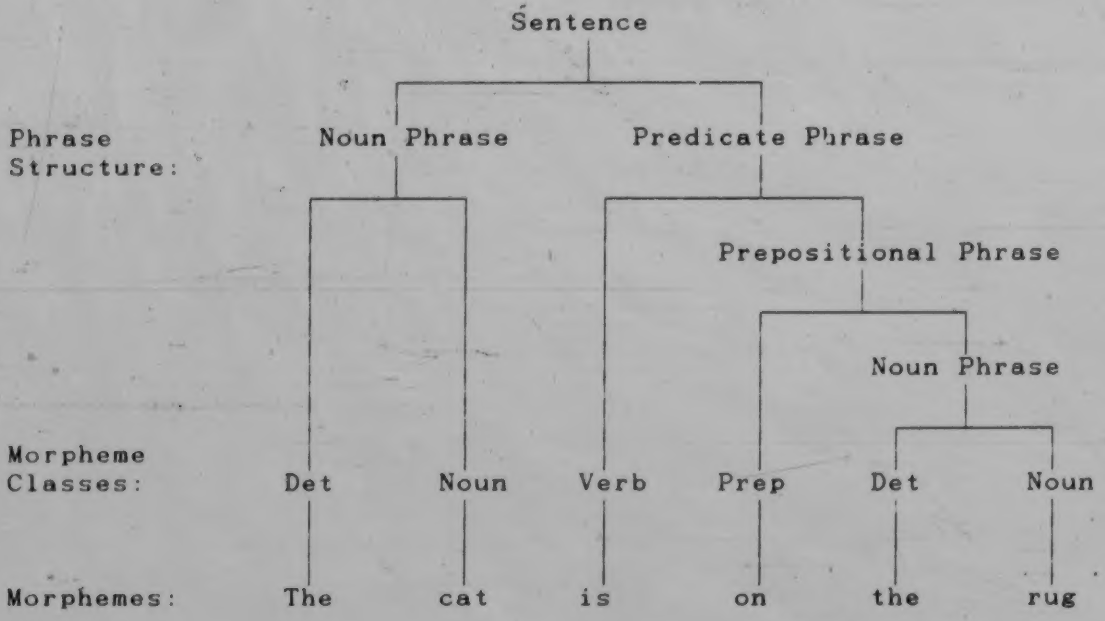
Once all the morpheme-classes and phrase-types have been identified, the linguist then must look for relations of inclusion between phrase-types. That is, the distributional analysis may indicate that a particular type of phrase, say a noun phrase, can occur within another phrase-type, say a subordinate clause. Furthermore, although the phrase types have been identified as sets of sequences of morphemes, the linguist should now be able to characterize the phrase types as sets of sequences of morpheme-classes, thus allowing a more compact representation of a phrase-type.

---

corpus entries when substituted into phonemic environments. If this empirical assumption is false, then Harris' additional semantic judgments will have to be introduced into the method.

When this study of inclusion relations is complete, the linguist is now able to syntactically transcribe the corpus. The syntactic transcription will not simply be a string, but rather, it will be a tree structure that indicates the relations of inclusion between morpheme-classes and the different levels of phrase-types. An example is given below:

Figure 2.2.5
Syntactic Transcription of an Utterance from the Corpus


Sentence
Phrase Structure:
Noun Phrase
Predicate Phrase
Prepositional Phrase
Noun Phrase
Morpheme Classes:
Det
Noun
Verb
Prep
Det
Noun
Morphemes:
The
cat
is
on
the
rug

Once all the utterances of the corpus have been transcribed in this manner, the linguist then looks for common syntactic patterns among the utterances. These common patterns are listed as the sentence-types of the language. This list of sentence-types, together with the list of morpheme-types and phrase-types, constitutes a description of the sentence-syntax of the language.[72]

Whorf and Bloomfield's program

As we have seen in Chapter One, Whorf was not satisfied with the way in which the structuralists viewed semantics. In order to formulate his Linguistic Relativity Hypothesis, Whorf required a semantic analysis that would allow him to compare different languages. According to Bloomfield's program, such a semantic analysis would eventually be available, but not for the foreseeable future. Bloomfield stated that:

> ... the statement of meanings is...the weak point in language study, and will remain so until the state of knowledge advances very far beyond its present state.[73]

For now, we can only make very basic semantic judgments, judgments of significance and synonymy, and the only reason for making these judgments is that they are necessary to effect a syntactic analysis.

---

[72] This presentation of discovery procedures for sentence-syntax is based on Z. S. Harris, "From Morpheme to Utterance", Language, 1946, 22, 169-180, reprinted M. Joos, ed., Readings in Linguistics I, (University of Chicago Press, 1957), and on Z. S. Harris, Structural Linguistics, (University of Chicago Press, 1951), chapters 15-19.

The idea that phrase types have inclusion relations is clearly explained in Harris, 1951, p. 332. A fuller discussion can be found in R. S. Wells, "Immediate Constituents", Language, 1947, 23, 81-117. Reprinted M. Joos, ed., Readings in Linguistics I, (University of Chicago Press, 1957).

The idea that sentence-syntax can be described by listing a series of sentence types is exemplified in Harris, 1946, and at a more elementary level in N. C. Stageberg, An Introductory English Grammar, (Holt, Rinehart and Winston, 1965), chapter 14.

[73] L. Bloomfield, Language, (1933, reprint University of Chicago Press, 1984), 146.

Whorf rejected this negative pronouncement on the current prospects for semantics. The following passage has already been quoted in Chapter One, but it is worth repeating here.

> At the beginning of investigation of a language, the "functional" type of definition, e.g. that a word of a certain class, say a "noun", is "a word which does so-and-so," is to be avoided when this is the ONLY test of distinction applied; for people's conceptions of what a given word "does" in an unfamiliar language may be as diverse as their own native languages, linguistic educations and philosophical predilections. The categories studied in a grammar are those recognizable through facts of a configurational sort, and these facts are the same for all observers. Yet I do not share the complete distrust of functional definitions which a few modern grammarians seem to show. After categories have been outlined according to configurational facts, it may be desirable to employ functional or operational symbolism as the investigation proceeds. Linked with configurative data, operational descriptions become valid as possible ways of stating the MEANING of the forms, "meaning" in such cases being a characterization which succinctly accounts for all the semantic and configurational facts, known or predictable.[74]

The problem with Whorf's position is that he gives us no idea as to what he means by "operational descriptions" that supposedly allow us to "succinctly account for all the semantic ... facts". So rather than focusing on Whorf's position, I would like to briefly consider the following question: If we were to follow Bloomfield's advice and wait for a general advancement of science such that we are in a position to formulate semantic analyses of natural languages, might those future semantic analyses be consistent with Whorf's Linguistic Relativity Hypothesis?

---

[74] B. L. Whorf, Language, Thought and Reality, edited by J. B. Carroll, (The M.I.T. Press, 1956), 88.

The answer to this question is negative. The question presupposes that Bloomfield's approach to semantics is correct, but this presuppostion must be examined. Consider again the following famous passage from Bloomfield's Language:

> The study of speech sounds without regard to meanings is an abstraction: in actual use, speech-sounds are uttered as signals. We have defined the meaning of a linguistic form as the situation in which the speaker utters it and the response which it calls forth in the hearer. The speaker's situation and the hearer's response are closely coordinated, thanks to the circumstance that every one of us learn to act indifferently as a speaker or as a hearer. In the causal sequence
>
> speaker's situation --> speech --> hearer's response,
>
> the speaker's situation, as the earlier term, will usually present a simpler aspect than the hearer's response; therefore we usually discuss and define meanings in terms of a speaker's stimulus.
>
> The situations which prompt people to utter speech, include every object and happening in their universe. In order to give a scientifically accurate definition of meaning for every form ['form' is Bloomfield's term for words, phrases, etc.] of a language, we should have to have a scientifically accurate knowledge of everything in the speaker's world. The actual extent of human knowledge is very small, compared to this. We can define the meaning of a speech-form accurately when this meaning has to do with some matter of which we possess scientific knowledge. We can define the names of minerals, for example, in terms of chemistry and mineralogy, as when we say that the ordinary meaning of the English word salt is 'sodium chloride (NaCl),' and we can define the names of plants or animals by means of the technical terms of botany or zoology., but we have no precise way of defining words like love or hate, which concern situations that have not been accurately classified - and these latter are the great majority.[75]

This passage is highly ambiguous as to exactly how we are to construe the meaning of a linguistic utterance. It seems possible to interpret what Bloomfield is saying in at least three distinct ways. He may be advocating behavioral semantics, in which the meaning of a form is given by the human behaviors (including perception) associated with its use; he may be advocating referential semantics, in which the meaning of a form is given by identifying the object or objects in the world which the form refers to or somehow describes; or he may be advocating translational semantics, in which the meaning of a form is given by translating it into some canonical language, specifically, the language of advanced science.

---

[75] L. Bloomfield, Language, (1933, reprint University of Chicago Press, 1984), 139.

These three alternatives do not necessarily sit well together. For example, if we take the meaning of a form as having something to do with behavior, or more exactly stimulus and response, then it is difficult to see why it is necessary to have tremendous scientific advances in all the sciences prior to doing semantics. All we really require is advances in behavioral psychology. So long as we can accurately describe the stimuli and responses of someone who uses the term 'love', what does it matter that we cannot give a scientific description of love? Alternatively, if we assume that the meaning of a form is basically a matter of reference, then again, it is not clear why scientific advancement is required. If 'salt' refers to that white stuff, then it refers to that white stuff no matter what my state of scientific advancement. The sentence "'Salt' means that white stuff" states exactly the same relation as "'Salt' means NaCl".[76]

The least objectionable way of weaving together the three threads in Bloomfield's semantics is to assume that the meaning of a form is primarily a matter of how it is translated in advanced scientific language. Then Bloomfield's behavioral considerations can be understood as evidence of which translation is correct. The referential theme can be understood as an allusion to the objects and events that are the causes of the stimuli associated with the use of linguistic forms.

I maintain, then, that the most charitable view of Bloomfield's semantics is that it is intended as a translational semantics, supported by behavioral evidence. Such an approach to semantic theorizing has been given a vigorous and sophisticated defense by W. V. O. Quine.[77] The object of semantic theorizing, according to Quine, is to produce a translation manual from the object language to a previously understood language. In order to be as objective as possible, the semantic theorist uses a variety of techniques that Quine collectively calls "radical translation." Quine does not demand that the target language of the translation be the language of advanced science, although he does require that it be properly "regimented", that is, cleared of many of the vagaries and ambiguities associated with unregimented natural languages.

---

[76] This is stated very loosely. It is assumed that the noun phrase 'that white stuff' that is employed in the meaning-giving sentence "'Salt' means that white stuff" is being used to identify a natural kind by pointing out a prototype of the natural kind.

[77] See especially W. V. O. Quine, Word and Object, (The M.I.T. Press, 1960), chapter 2.

In reading Quine's work on natural language semantics, one is struck by the similarity of tone to the writings of Bloomfield and Harris. Quine shares the structuralist linguists' disdain of the hasty conclusion. This unwillingness to generalize beyond the behavioral evidence led Quine to formulate one of his most famous doctrines, the thesis of the indeterminacy of translation. The idea is that the behavioral evidence available to the field linguist will always be consistent with multiple translation manuals that are different in non-trivial ways. In fact, Quine has linked this thesis to Whorf's thesis.

> One frequently hears it urged that deep differences in language carry with them ultimate differences in the way one thinks, or looks upon the world. I would urge that what is most generally involved is indeterminacy of correlation. There is less basis of comparison - less sense in saying what is good translation and what is bad - the further we get away from sentences with visibly direct conditioning to non-verbal stimuli and the farther we get off home ground.[78]

Quine's claim is that Whorf has mistaken a methodological problem, indeterminacy of translation, for a substantive issue, the Linguistic Relativity Hypothesis. According to Quine, the evidence that Whorf marshalls to demonstrate the Linguistic Relativity Hypothesis actually shows nothing of the kind; it merely displays the fact that it is possible to generate a rather unusual translation manual to correlate the native language with the previously understood target language. Many other alternative manuals would also be compatible with the evidence, according to Quine. However, Quine warns us that there is no basis for going beyond this observation to the conclusion that the speakers of the language actually have substantively different thought processes than we do.

---

[78] Ibid., 77-78.

Quine's transformation of Whorf's Linguistic Relativity Hypothesis from a substantive issue into a methodological issue would be a neat trick if we could accept his semantics. However, his semantics is seriously flawed. Like all translational approaches to semantics, it merely postpones the problem. Donald Davidson points this out in the following passage:

> I do not think that a translation manual is the best form for a theory of interpretation [Davidson's term for a semantic theory of a natural language] to take.
>
> When interpretation is our aim, a method of translation deals with a wrong topic, a relation between two languages, where what is wanted is an interpretation of one (in another, of course, but that goes without saying since any theory is in some language). We cannot without confusion count the language used in stating the theory as part of the subject matter of the theory unless we explicitly make it so. In the general case, a theory of translation involves three languages: the object language, the subject language, and the metalanguage (the languages from and into which translation proceeds, and the language of the theory, which says what expressions of the subject language translate which expressions of the object language). And in this general case, we can know which sentences of the subject language without knowing what any of the sentences of either language mean (in any sense, anyway, that would let someone who understood the theory interpret sentences of the object language). If the subject language happens to be identical with the language of the theory, then someone who understands the theory can no doubt use the translation manual to interpret alien utterances; but this is because he brings to bear two things he knows and that the theory does not state: the fact that the

> subject language is his own, and his knowledge of how to interpret utterances of his own language.[79]

Davidson's critique of translational semantics is based on the following tenet: Speakers of a natural language can understand (Davidson uses the term 'interpret') the sentences of their language. Therefore a reasonable criterion of adequacy for any semantic theory is that it should state information that is sufficient to permit one to understand the language that it is a semantic theory of. In other words, a semantic theory is adequate only if comprehension of that theory will result in an understanding of the language that it is a theory of. Davidson points out that a translational approach to semantics completely fails to meet this minimal requirement.

The line of argument that we have been pursuing goes like this: Whorf's Linguistic Relativity Hypothesis is only as clear as his explication of semantic phenomena. However, Whorf was extremely vague about semantics. He approaches the subject initially from a Bloomfieldian perspective, recommending a bolder approach, but it is not at all clear what he is getting at. Instead of guessing at Whorf's intentions we asked another question: Would the eventual development of semantic theory as characterized by Bloomfield be consistent with the Linguistic Relativity Hypothesis? Bloomfield's approach to semantics is most consistently understood as advocating a translational approach to semantics based on behavioral evidence. Such an approach has been worked out in detail by Quine. Quine has pointed out that his approach shows that the phenomena that led Whorf to formulate the Linguistic Relativity Hypothesis is actually a methodological artifact rather than an indication of a substantive psychological phenomenon. However, Quine's dismissal of the Linguistic Relativity Hypothesis is premature, because his translational approach to semantics does not meet a basic criterion of adequacy for a semantic theory.

---

[79] D. Davidson, "Radical Interpretation", Dialectica, 1973, 27, 309-323. Reprinted in D. Davidson, Inquiries into Truth and Interpretation, (Oxford University Press, 1984), 129.

Our results so far are negative. Bloomfield's approach to semantics will not support the Linguistic Relativity Hypothesis. Whorf's own "liberalization" of Bloomfieldian semantics is too vague to take us anywhere. Consequently in section 2.3 an entirely different approach to semantic theorizing will be considered, and it will be shown that much of Whorf's writing is very similar in spirit to this approach. It will then be asked if this approach is defensible, and if so, whether it is consistent with the Linguistic Relativity Hypothesis.

However, before turning to this alternate approach, the fate of the structuralist program will be briefly reviewed. As will be shown, Noam Chomsky presented a strikingly different approach to linguistic theorizing. Many of Chomsky's insights have been incorporated into the semantic theories to be considered later, so we will do well to consider them now.

Chomsky's critique of Bloomfield's program

Noam Chomsky is responsible for an extremely interesting and influential argument against the use of discovery procedures. In order to fully understand Chomsky's argument, it is necessary to understand the basic concepts of mathematical linguistics, a field that Chomsky largely initiated.

Mathematical Linguistics

Mathematical linguistics is a branch of mathematics that can be construed as a formalization of that part of the study of language that we have been calling sentence-syntax. In what follows, the term sentence-syntax will be understood as a type of mathematical object.

A sentence-syntax, then, is defined as a triple $SS = \{V_N, V_T, P\}$ consisting of a non-terminal vocabulary V_N, a terminal vocabulary V_T, and a set of productions P, with the following properties:

(1) V_N, V_T and P are finite non-empty sets

(2) $V_N \cap V_T = \emptyset$, the null set

(3) $S \in V_N$, where S is the start-symbol

(4) P is a subset of $V^+ \times V^*$, where V^* is the union of V_T and V_N, and V^+ is V^* with the null-string excluded.[80]

Given this definition of a sentence syntax, we can define a sentence as any $s \in V^*$ such that $S \Rightarrow\!\!> s$, where the symbol '=>>' is explained as follows: The productions P (sometimes called production rules) are pairs of strings, conventionally written in the form $a \rightarrow b$, where $a \in V^+$ and $b \in V^*$. Each production is to be interpreted as a rewrite rule; that is, the production above says that a can be rewritten as b. Such rules apply in any context, so by applying the production just given we can rewrite

---

[80] W. Levelt, Formal Grammars, (Mouton, 1974), Vol 1, 5.

'xay' as 'xby'. This can be symbolized as xay => xby, where the symbol '=>' is read as "directly derives". If there is a set of productions a(1) => a(2), a(2) => a(3), . . . a(n-1) => a(n), then we can write a(1) =>> a(n), which is read "a(1) derives a(n)". Thus the definition of a sentence (with respect to a particular sentence-syntax) as any $s \in V^*$ such that S =>> s means that a sentence is any string that can be derived from the start symbol by means of the production rules.

The following is a simple example of a sentence-syntax:

Figure 2.2.6
A Simple Example of a Sentence-Syntax

V_T = { the_dog, the_cat, the_mouse, runs, sleeps, eats, frightens, and, or }

V_N = { S, PROPER_NOUN, TRANSITIVE_VERB, INTRANSITIVE_VERB, CONNECTIVE }

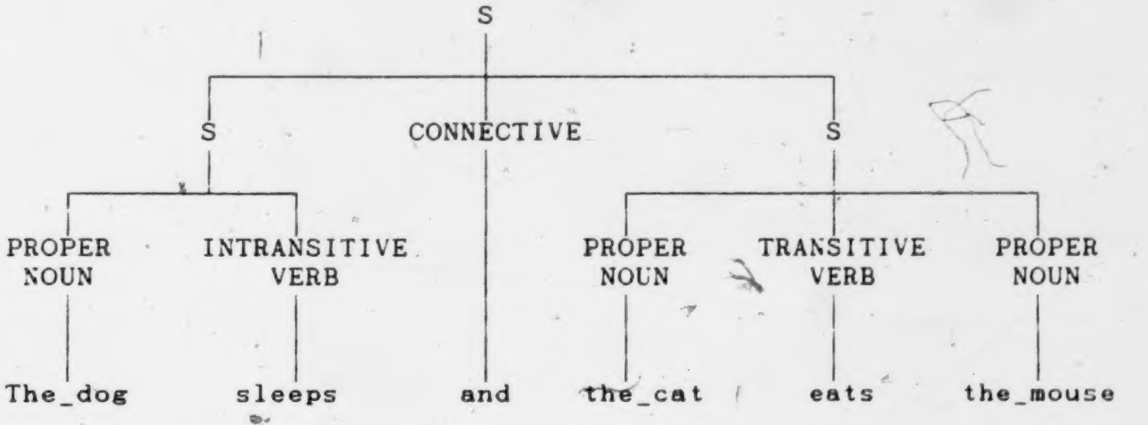
P = { S -> PROPER_NOUN + TRANSITIVE_VERB, (1)
S -> PROPER_NOUN + INTRANSITIVE_VERB + PROPER_NOUN, (2)
S -> S + CONNECTIVE + S, (3)
PROPER_NOUN -> the_dog, (4)
PROPER_NOUN -> the_cat, (5)
PROPER_NOUN -> the_mouse, (6)
INTRANSITIVE_VERB -> runs, (7)
INTRANSITIVE_VERB -> sleeps, (8)
TRANSITIVE_VERB -> eats, (9)
TRANSITIVE_VERB -> frightens, (10)
CONNECTIVE -> and, (11)
CONNECTIVE -> or } (12)

NOTE: '+' is used to indicate concatenation.

This syntax can be used to generate the sentence 'The_dog sleeps' as follows. Start with 'S', the start symbol. Use production (1) to derive 'PROPER_NOUN + TRANSITIVE_VERB', use production (4) to derive 'The_dog + TRANSITIVE_VERB', use production (8) to derive 'The_dog + sleeps', and then drop the meta_linguistic concatenation sign to obtain 'The_dog sleeps'.

From this example we can see what motivates the distinction between V_N, the non-terminal vocabulary, and V_T, the terminal vocabulary. By definition, the sentences produced by the syntax are composed entirely of terminal elements. The non-terminal elements, on the other hand, correspond to grammatical categories, i.e., nouns, verbs, sentences, etc. These grammatical categories are employed in the production rules in the generation of particular sentences. Indeed, we can represent the derivation of any sentence from this syntax by means of a "phrase-structure diagram" which will indicate the role that the non-terminal vocabulary had in the particular derivation. For example:

Figure 2.2.7
Phrase Structure Diagram


S
S
CONNECTIVE
S
PROPER NOUN
INTRANSITIVE VERB
PROPER NOUN
TRANSITIVE VERB
PROPER NOUN
The_dog
sleeps
and
the_cat
eats
the_mouse

4

(However, it should be noted that not every sentence-syntax will produce derivations that can be captured with phrase structure diagrams. This will become apparent below.)

It is noteworthy that the example sentence-syntax is "recursive", and is therefore capable of generating an infinite number of sentences from a finite number of rules. Specifically, note the third production rule: S -> S + CONNECTIVE + S. This is called a recursive rule because one of the elements, viz. S, appears on both sides of the rewrite sign. Since there is no limit to the number of times a production rule can be applied, this means that there is no limit to the length of sentence that can be derived from the grammar. For example, applying this recursive rule allows the derivation of strings of the form 'The_dog sleeps and the_dog sleeps and the_dog sleeps and ...' for arbitrarily many repetitions of 'and the_dog sleeps'. Note that this means that the sentence-syntax can generate an infinite number of sentences, for even though each sentence has a finite length, there is no upper bound on this length (just as there are an infinite number of numerals, even though each numeral is composed of a finite number of symbols).

The example sentence-syntax given above generates derivations that can be represented by phrase structure diagrams. However, not every sentence-syntax has derivations that can be so represented. Sentence-syntaxes that "go beyond" phrase structure are, in a sense to be made precise, more powerful that phrase structure sentence-syntaxes. Chomsky has provided us with a precise way of stating these points. His system is called the

"Chomsky Hierarchy of Grammars", which in our terminology becomes the "Chomsky Hierarchy of Sentence-Syntaxes". The hierarchy is as follows:

A Type-0 Sentence-Syntax is defined exactly the same way that a sentence-syntax is defined. Thus, every sentence-syntax is a Type-0 sentence-syntax.

A Type-1 Sentence-Syntax contains only productions where the number of symbols on the right hand side of the rewrite symbol is always greater than or equal to the number of symbols on the left hand side of the rewrite symbol. Consequently, with a Type-1 sentence-syntax it is never possible to reduce string length as a result of applying productions.

A Type-2 Sentence-Syntax contains only productions of the form A -> B where (1) A consists of only one member of V_N, and (2) B $\in V^+$, i.e., B is not the null-string.

A Type-3 Sentence-Syntax contains only productions of the form A -> B or A -> bB where (1) A and B are single members of V_N, and (2) b is a single member of V_T.

How would the example sentence-syntax specified in figure 2.2.6 be classified in terms of this hierarchy? First, it is a Type-0 sentence-syntax, obviously, since every sentence-syntax is Type-0. It is also Type-1, since none of the productions result in a reduction in string length. It is also Type-2 since every production has a left hand side that is a single member of V_N, and no right hand side is the null string. However, it is not Type-3 because the first three productions fail to conform to the Type-3 restrictions.

It can be seen, then, that these types form a hierarchy of nested sets, that is, Type-3 syntaxes are a subset of Type-2s, which are a subset of Type-1s, which are a subset of Type-0s.

We can now introduce the following terminology.

An Unrestricted Rewrite Syntax is any member of the set of syntaxes obtained by set subtracting the Type-1s from the Type-0s.

A Context-Sensitive Syntax is any member of the set of syntaxes obtained by set subtracting the Type-2s from the Type-1s.

A Context-Free Syntax is any member of the set of syntaxes obtained by set subtracting the Type-3s from the Type-2s.

A Regular Syntax is any member of the set of Type-3 syntaxes.

A final bit of terminology:

An Unrestricted Language is a language that can be generated only by sentence-syntaxes that are unrestricted rewrite systems.

A Context-Sensitive Language is a language that can be generated by some context-sensitive phrase structure syntax, but cannot be generated by any context-free phrase structure syntax, or by any regular syntax.

A Context-Free Language is a language that can be generated by some context-free phrase structure syntax, but cannot be generated by any regular syntax.

Finally, a Regular Language is a language that can be generated by some regular syntax.

With this background in the basic concepts of mathematical linguistics we can reconstruct Chomsky's case against structuralist linguistics. It consists of two claims:

(1) Structuralist linguistics in effect endorses a limit on the power of a syntax, namely, that context-sensitive phrase structure grammars are the most powerful that the techniques will permit.

(2) However, the natural languages used by humans are unrestricted languages. Consequently, structuralist linguistics is inherently incapable of providing a correct account of human natural languages.

These two claims will now be examined.

(1) Chomsky on the Discovery Procedures

Our examination of the discovery procedures for sentence-syntax has shown that the structuralist linguist tries to formulate a set of morpheme-classes and a set of phrase-types. Phrase-types can be represented as sequences of morpheme-classes, but some phrase-types can be viewed as sequences of other phrase-types. There are relations of inclusion between phrase-types. What the structuralist discovery procedures can unfold, then, is a hierarchical arrangement of phrase-types. Armed with a specification of the phrase-types of a language, each sentence of the language can be given an immediate-constituent analysis. That is, each sentence can be viewed as a sequence of phrase-types, each of which in turn can be viewed as a sequence of phrase-types or ultimately a sequence of morpheme-classes.

In his classic Syntactic Structures, Chomsky links this approach to syntactic theory to the mathematical theory of linguistics:

> Customarily, linguistic description on the syntactic level is formulated in terms of constituent analysis (parsing). We now ask what form of grammar [i.e., formal sentence-syntax] is presupposed by description of this sort. ...
>
> Each such grammar is defined by a finite set E of initial strings and a finite set F of 'instruction formulas' of the form X -> Y interpreted: 'rewrite X as Y.' Though X need not be a single symbol, only a single symbol of X can be rewritten in forming Y.[81]

Careful examination of the last paragraph will convince the reader that the sentence-syntax discovery procedures described above will in effect generate exactly the "instruction formulas" (or "productions", as we have been calling them) that Chomsky describes. Furthermore, it should be clear from the last sentence of the quotation that the most powerful sentence-syntax that can be characterized by "instruction formulas" of this type will be a context-sensitive sentence-syntax. However, the quotation limits us to that class of context-sensitive syntaxes that generate hierarchical tree structures. These are sometimes called context-sensitive phrase structure syntaxes.[82] Consequently, Chomsky concludes that the discovery procedures, in effect, rule out unrestricted rewrite systems and certain types of context-sensitive syntaxes (specifically, those that produce derivations that cannot be represented with phrase structure diagrams) as formalizations of natural languages.[83]

---

[81] N. Chomsky, Syntactic Structures, (Mouton, 1972), 26, 29.

[82] As in W. J. M. Levelt, Formal Grammars in Linguistics and Psycholinguistics, Volume II: Applications in Linguistic Theory, (Mouton, 1974), 36-37.

(2) **Chomsky on Human Languages**

Chomsky and his followers have argued that mathematical linguistics provides a formalization for the development of syntactic descriptions of natural languages. Furthermore, they have argued natural languages require a syntax that is very "high" in the Chomsky hierarchy of syntaxes. Specifically, they have argued that the appropriate level of syntax for human languages is an unrestricted rewrite system.

There are two ways in which a formal syntax can be viewed as "adequate" as a formalization of the syntactic structure of a natural language. A formal syntax is "observationally adequate" if it generates all and only the sentences of the natural language. If, in addition, the derivations of sentences are somehow consonant with our intuitions (for example, if the derivation of a particular sentence assigns a phrase structure to the sentence, and that phrase structure conforms with our intuition of how the sentence is "parsed") then we can say that the syntax is "descriptively adequate".[84] Chomsky and his followers have argued that neither regular sentence-syntaxes nor context-free sentence-syntaxes are observationally adequate for various natural languages. They have also argued that only unrestricted rewrite systems are descriptively adequate for natural languages.

---

[83] A more thorough demonstration of the phrase structure character of structuralist linguists can be found in P. Postal, "Limitations of Phrase Structure Grammars", in The Structure of Language, edited by J. A. Fodor and J. H. Katz, (Prentice-Hall, 1964).

[84] These definitions are taken from W. J. M. Levelt, Formal Grammars in Linguistics and Psycholinguistics, Volume II: Applications in Linguistic Theory, (Mouton, 1974), 8. I find Levelt's terms more precise and more useful than the well-known distinction that Chomsky makes between "descriptive adequacy" and "explanatory adequacy" in N. Chomsky, Aspects of the Theory of Syntax, (The M.I.T. Press, 1965).

First let us consider the arguments regarding observational adequacy. In Syntactic Structures, Chomsky presented a sketch of a "proof" that English is not a regular language by considering the property of "self-embedding".[85] A language is said to be self-embedding if every syntax that generates that language is self-embedding. A syntax is self-embedding if there is a variable $B \in V_N$, and elements a, $b \in V^+$ such that there is a production $B \Rightarrow aBb$.[86] It is a consequence of this definition that a regular syntax cannot be self-embedding since the productions in a regular grammar can never take the form prescribed above. Chomsky claims that English cannot be a regular language because it is a self-embedding language.

The evidence that Chomsky marshals for the self-embeddedness of English is the following:

> Let S_1, S_2, S_3, ... be declarative sentences in English. Then we can have such English sentences as:
>
> (11) (i) If S_1, then S_2.
> (ii) Either S_3, or S_4.

---

[85] N. Chomsky, Syntactic Structures, (Mouton, 1972), 21-23. Chomsky's detailed argument is found in N. Chomsky, "Three Models for the Description of Language", I.R.E. Transactions on Information Theory, 1956, vol. IT-2, proceedings of the symposium on information theory.

[86] W. J. M. Levelt, Formal Grammars in Linguistics and Psycholinguistics, Volume I: An Introduction to the Theory of Formal Languages and Automata, (Mouton, 1974), 21.

> (iii) The man who said that S_5, is arriving today.
>
> In (11i), we cannot have 'or' in place of 'then'; in (11ii), we cannot have 'are' instead of 'is'. In each of these cases there is a dependency between words on opposite sides of the comma (i.e., 'if'-'then', 'either'-'or', 'man'-'is'). But between the interdependent words, in each case, we can insert a declarative sentence S_1, S_3, S_5, and this declarative sentence may in fact be one of (11i-iii). Thus if in (11i) we take S_1 as (11ii) and S_3 as (11iii), we will have the sentence:
>
> (12) if, either (11iii), or S_4, then S_2,
>
> and S_5 in (11iii) may again be one of the sentences of (11). It is clear, then, that in English we can find a sequence a + S_1 + b, where there is a dependency between a and b, and we can select as S_1 another sequence containing c + S_2 + d, where there is another dependency between c and d, then select as S_2 another sequence of this form, etc. A set of sentences that is constructed this way (and we see from (11) that there are several possibilities available for such a construction - (11) comes nowhere near exhausting these possibilities) will have all of the mirror image [Chomsky's term for self-embeddedness] properties ... [which are exclusive of] ... the set of finite state languages [Chomsky's term for regular languages]. Thus we can find various kinds of non-finite state [i.e., non-regular] models within English.[87]

However, merely to point out that a non-regular "model" can be found in English does not prove that English is a regular language. Consider the regular syntax that is defined by the following productions:

```
S -> S + W
S -> W
W -> any word of English
```

This syntax will generate "sentences" of any length, consisting of any combination of English words. Included in this output will be the non-regular "models" of which Chomsky writes. But clearly, the existence of these models is not inconsistent with the fact that the language was generated by a regular syntax. The existence of a non-regular model points to a non-regular syntax only if other conditions are met. For example, an additional condition would be the following:

---

[87] N. Chomsky, Syntactic Structures, (Mouton, 1972), 22-23.

> All sentences of the form
>
> $\text{If}^n\ S_1\ (\text{then}\ S_2)^m$
>
> are ungrammatical when n is not equal to m.

If this condition is true, and moreover, if there are similar conditions stated for all the other sources of self-embedded sentences that Chomsky has referred to, and moreover, these conditions are true, then this can be used a the basis of a formal proof that English is not a regular language.[88] However, it is easy to give counter-examples that show that these conditions do not hold in English. For example, 'If it rains, it pours' is an acceptable English sentence even though the number of 'if's does not equal the number of 'then's.

Consequently Chomsky's "proof" that English is a regular language, and therefore that any regular syntax of English is observationally inadequate, fails. Furthermore, the argument schema used in Syntactic Structures did not require that the extra conditions had to be met. A quick reading gives one the impression that merely to point out the existence of an embedded model is to demonstrate English is not a regular language.[89] This is, of course, not true.

---

[88] For more details see W. J. M. Levelt, Formal Grammars in Linguistics and Psycholinguistics, Volume II: Applications in Linguistic Theory, (Mouton, 1974), 22-26.

This faulty argument schema was subsequently used by a number of other linguists, including Paul Postal, who produced a highly influential "proof" that Mohawk (a North American Indian language) is not context-free.[90] These arguments have been widely accepted as valid by many linguists working in the post-Chomskyan era. However, it appears that all of these arguments either use the faulty argument schema of Syntactic Structures, or, if they do invoke the extra conditions required to permit the proof to go through, the extra conditions are empirically false.[91]

In conclusion, Chomsky and a number of his followers tried to formally prove that low-level syntaxes such as regular syntaxes and phrase structure syntaxes were "observationally inadequate" to describe natural languages. However, all of these "proofs" failed.

---

[89] I am not suggesting that Chomsky is not aware that the proof requires the extra conditions. To be fair, on page 23 of Syntactic Structures (Mouton, 1957), he notes that such extra conditions are required, and he discusses them in detail in N. Chomsky, "Three Models for the Description of Language", I.R.E. Transactions on Information Theory, 1956, vol, IT-2, proceedings of the symposium on information theory. However, he does present a misleading and incomplete argument in Syntactic Structures, and this misleading argument was widely accepted as valid. For an example this bad influence, see J. Lyons, Chomsky, (Fontana/Collins, 1970), 54.

[90] P. Postal, Constituent Structure: A Study of Contemporary Models of Syntactic Description, (Indiana University Press, 1964).

[91] For a detailed analysis of the inadequacy of such arguments see R. T. Daly, Applications of the Mathematical Theory of Linguistics, (Mouton, 1974), chapters 3 and 4.

However, Chomsky and his followers also argued that low-level syntaxes are also "descriptively inadequate", that is, they fail to conform to our intuitions about linguistic structure. These arguments do not have the problems associated with the observational adequacy arguments, but they are also less satisfying, in that they are only as strong as the intuitions on which they are based.

Chomsky argued that a natural language syntax should have a phrase structure base and an additional transformational component. The phrase structure base is either a context-free syntax or a context-sensitive phrase structure syntax. The transformational component consists of productions that operate on the output of the phrase structure base. These transformations are not restricted in the way that context-free or context-sensitive productions are (see the definitions of Type-1 and Type-2 syntaxes above). For example, some of the transformations that have been proposed result in the deletion of elements of a string. Obviously, any syntax containing such transformations is an unrestricted rewrite system.

Chomsky and his followers have developed a large number of arguments to demonstrate that transformational syntaxes are superior to phrase structure syntaxes from the perspective of descriptive adequacy. One of the earliest and most influential of such arguments is Chomsky's discussion of active and passive sentences.[92] Chomsky pointed out that any phrase structure syntax has to propose separate derivations for sentence pairs such as 'John eats lunch' and 'Lunch is eaten by John'. A transformational syntax, on the other hand, can have phrase structure productions that generate only one of the pair of sentences, while the other is derived from the first by a transformational rule. From a "descriptive adequacy" perspective the transformational syntax is superior, for it "accounts" for the intuitive connection between active-passive pairs in a way that a phrase structure grammar cannot.

---

[92] N. Chomsky, Syntactic Structures, (Mouton, 1972), 42-43.

So although Chomsky and his followers failed to prove the observational inadequacy of "lower" syntaxes, they did succeed in making the softer argument that these lower syntaxes are descriptively inadequate. These arguments were very influential, so that it is not uncommon nor inappropriate to say that Chomsky affected a "revolution" in linguistics.[93] This revolution resulted in fundamental changes in the goals and methods of linguistics as formulated by Bloomfield. Two of these changes play a role in the sections to follow, and therefore warrant explanation.

1. The Methodology of Linguistics

Chapter 6 of Syntactic Structures is called "On the Goals of Linguistic Theory". In this chapter Chomsky characterized the structuralist discovery procedures as an attempt to specify a mechanism that takes a corpus as input and generates a syntax as output. Chomsky contrasts this with an alternate view: A syntax should be regarded as a theory; a theory of the syntactic structure of the language in question. No one (any longer) thinks that there is a strictly mechanical procedure for generating theories in physics from physical evidence, and similarly, we should no longer think that a syntax can be mechanically generated from a corpus. Instead, syntaxes should be developed in the same way as theories, that is, as speculative accounts. Then, of course, the theories must be put to the test. So the proper view of linguistic methodology is this: We have a corpus and a number of speculative theories. The purpose of linguistic methodology is to allow us to evaluate and rank the theories in terms of their "fit" with the corpus. Linguistic methodology cannot spare us the creative task of generating theories. Nor can it identify the "one true" theory; it can, at best, provide us with a ranking of theories. According to Chomsky, the methods of linguistics are the same as for any other science attempting to establish general theories.

---

[93] See J. Searle, "Chomsky's Revolution in Linguistics", The New York Review of Books, 1972, reprinted in On Noam Chomsky: Critical Essays, edited by G. Harman, (Doubleday Anchor, 1974).

Although Chomsky's general point is inherently plausible, his detailed arguments in favor of viewing syntaxes as theories were not overwhelming. In opposition to Chomsky, one could make a case that transformational productions could be determined by merely extending structuralist discovery techniques. In fact, the structuralist Zelig Harris actually proposed transformations in 1957 within the context of a structuralist paradigm, and even outlined the discovery procedures that he felt were required to identify transformations.[94]

---

[94] Z. S. Harris, "Co-occurrence and Transformation in Linguistic Structure", Language, 1957, 33, 283-340. Excerpts

However, it has subsequently been clarified that there are definite limits on discovery procedures. In 1967 E. M. Gold published a paper that proved that certain syntaxes cannot be derived from a corpus by any mechanical procedure.[95] Gold considers a algorithm that accepts as sentences from a corpus as input, one at a time. As each sentence is accepted, the algorithm produces a syntax that is compatible with the input so far. If after a finite number of inputs the output stabilizes, then we say that the syntax of the language on which is corpus is based is "learnable". A class of languages (e.g., the class of context-free languages) is learnable if every language in that class is learnable.

Gold showed that if the input consists of a sequence of sentences that are grammatical in the language (and no other information is provided), then only finite languages (i.e., languages that contain no recursive productions) are learnable. On the other hand, if the input consists of both grammatical and ungrammatical strings, together with a judgement on the grammaticality of the string, then a subset of the unrestricted rewrite languages will not be learnable; however, all other language classes will be.

Gold's work can not be directly applied to the discovery procedures of the structuralists, since neither of his

---

reprinted in Syntactic Theory I: Structuralist, edited by F. W. Householder, (Penguin, 1972), 156.

[95] E. M. Gold, "Language Identification in the Limit", Information and Control, 1967, 6, 441-474. Gold's work is discussed in W. J. M. Levelt, Formal Grammars in Linguistics and Psycholinguistics, Volume I: An Introduction to the Theory of Formal Languages and Automata, (Mouton, 1974), 121-124.

characterizations of the input to the syntax-inferring algorithm contain the synonymy judgments that are integral to the structuralist's procedures. However, his work does definitely refute the easy confidence that the structuralists had in deriving syntaxes from the data in a mechanical fashion. Together with Chomsky's suggestion that the relation of syntax to corpus should be no different that the relation to theory and fact in other sciences, Gold's demonstration of the inherent limitations of syntax-inferring algorithms definitely suggests that discovery procedures represent a very questionable method, and should therefore be abandoned.

Indeed, they have been abandoned by the majority of linguists, and a more openly speculative approach to theorizing has characterized linguistics for the past three decades.[96]

---

[96] I have argued that the demise of the discovery procedures was largely due to the introduction of transformations into linguistic theory. Some of these transformations are unrestricted rewrite rules and cannot be accommodated in any phrase structure grammar. In recent years there has been a reaction against transformation rules in natural language syntax. Transformations play a less important role in Chomsky's current "Government-Binding" theory. Even more significantly, some theorists have suggested that transformations are not necessary at all. For example, in G. Gazdar, et. al., Generalized Phrase Structure Grammar, (Harvard University Press, 1985), the authors argue that a suitably generalized phrase structure syntax is adequate for human natural languages (although in footnote 7 on page 16 they express some reservations). However, it is clear that these authors are not advocating a return to the discovery procedures. They are offering a very abstract and highly formalized speculative theory in the same spirit that the transformationalists have presented their theories in the past several decades. It is very unlikely that the clock will ever be turned back to the methods of the structuralists, even if all the theories of the transformationalists are ultimately rejected.

2. Linguistics as Cognitive Psychology

Chomsky has argued that just as the structuralist linguist cannot simply derive the syntax from a corpus, neither can the child learning a language simply derive his grasp of syntax from exposure to a corpus using the mechanisms of learning postulated by behavioristic psychology. Chomsky claims that the evidence to which the child is exposed is of such poor quality, and the syntax that he eventually learns is of such complexity, that the acquisition of the syntax could not possibly be explained on the assumption that the child has only primitive learning mechanisms at his disposal.

To help make Chomsky's point, let us make some suppositions. Suppose that we have determined that human natural languages are all unrestricted rewrite systems. Suppose also that we have formalized behavioristic learning principles as an algorithm. Suppose also that we have formalized the child's exposure to language as the input to the algorithm. Then it is conceivable that someone could prove a theorem, in the manner of Gold's work, that it is impossible for a child to learn a human language on the basis of the evidence available to the child. Assume for argument that such a theorem has been proven. But this would contradict the obvious fact that children do learn languages. Therefore one of the assumptions that went into the proof must be wrong. Chomsky would hold that the incorrect assumption is that children have primitive learning mechanisms. Since children do learn highly complex (unrestricted rewrite) languages, they must have a learning algorithm that, in effect, rules out certain subsets of syntaxes. That is, by only entertaining a restricted hypothesis space of unrestricted rewrite languages, computational problem is reduced in complexity, and the child is capable of learning a highly complex language on the basis of minimal input. (Of course, there is no such theorem. However, I think it fair to say that Chomsky would hold that such developments are highly plausible.)

To say that the child is born with a learning mechanism that will only entertain a restricted hypothesis space of syntaxes is not much different from saying that the child has innate syntactic knowledge, and Chomsky says this quite explicitly.[97] This leads to a new goal for linguistic theory: Linguistic theory in its most generalized form should aim at characterizing this innate syntactic knowledge that is common to all human beings. Generalized linguistic theory thus becomes a branch of cognitive psychology.[98]

The idea that linguistics is psychology in disguise puts even more distance between contemporary linguists and the structuralists. If a syntactic theory is viewed as a theory about the mind, then the structuralist's discovery procedures seem even less appropriate.

---

[97] N. Chomsky, "Recent Contributions to the Theory of Innate Ideas", Boston Studies in the Philosophy of Science, Volume III, (The Humanities Press, 1968). Reprinted in The Philosophy of Language, edited by J. R. Searle, (Oxford University Press, 1971).

[98] The best presentation of the idea that linguistics is a branch of cognitive psychology that I am aware of is N. Chomsky, Reflections on Language, (Pantheon, 1975), chapter 1. An explicit discussion of how this psychologized conception of linguistics affects the details of syntactic theory can be found in P. W. Culicover and K. Wexler, "Some Syntactic Implications of a Theory of Language Learnability", in Formal Syntax, edited by P. W. Culicover, T. Wasow, and A. Akmajian, (Academic Press, 1977).

Summary

Bloomfield developed an approach to the study of language that dominated American linguistics until it was eventually supplanted by Chomsky's transformational approach. Bloomfield argued that although semantic considerations are required in order to properly conceptualize what natural language is, detailed semantic studies must be put on hold indefinitely. For the present, linguists should concentrate on the syntactic description of language, and only a modicum of semantic judgements are required for that task. Syntactic descriptions should be derived from a corpus of utterances through the application of discovery procedures.

Whorf's Linguistic Relativity Hypothesis was formulated in the heyday of structuralist linguistics. Whorf was at odds with the thinking of his time, since the Linguistic Relativity Hypothesis requires a fairly thoroughly developed notion of semantic theory; something that was not available in structuralist linguistics. However, Whorf did not develop a clear conception of semantic theory himself. Furthermore, the Bloomfieldian conception of a semantics of the future grounded in the other sciences won't work for Whorf either. The problem is not that Bloomfieldian semantics is indefinitely delayed; rather, the problem is that Bloomfieldian semantics is, at bottom, merely a translation into another language. Translation does not meet a basic criterion that a semantic theory should meet: it should impart information sufficient to give the reader of the semantic theory an understanding of the object language of the theory.

Bloomfield's program seems inherently incapable of providing the semantic theory required to precisely formulate Whorf's Linguistic Relativity Hypothesis. This is just as well, since Chomsky's work in the late fifties cast grave doubts on the viability of that program. By rigorously formalizing the notion of a syntax, Chomsky argued that human languages are characterized by syntactic complexities that the structuralist's discovery procedures are inherently incapable of capturing. This led to fundamental changes in how syntaxes were viewed. Firstly, they began to be viewed as theories, and therefore they need not be "derived" from facts; rather, they can be developed in the arm-chair (or wherever) so long as they are evaluated against the facts. Secondly, linguistics began to be viewed as a branch of cognitive psychology.

As we shall see in the next section, these changes in how syntax was viewed contributed to a much more optimistic attitude toward the development of semantic theories. The semantic theories that developed in the wake of the Chomskyan revolution are, on the face of it, compatible with a number of Whorf's claims about semantics. We now turn to those theories to see if they can provide a semantic foundation for the Linguistic Relativity Hypothesis.

2.3 CONCEPTUAL SEMANTICS

As a result of Chomsky's critique of the Bloomfieldian program, the idea that a syntax could be mechanically generated from a corpus was rejected by most linguists. Syntactic theorizing became a respectable activity. In his first major work, Syntactic Structures, Chomsky continued to adhere to the Bloomfieldian principle that the syntax of a language could be specified independently of the language's semantics. However, by 1965, Chomsky had reversed his thinking. In Aspects of the Theory of Syntax, Chomsky held that a semantic theory should be developed in concert with a syntactic theory to provide an overall specification of the grammar of the language.

One reason for the introduction of semantics was simply a more liberal attitude toward theorizing. Once Chomsky had broken down the Bloomfieldian barriers against syntactic theorizing, there was also less resistance to semantic theorizing. Another reason was that as transformational syntax evolved, puzzles began to emerge that seemed to naturally call for a resolution that seemed more semantic than syntactic. It appeared to many that a combined syntactic and semantic theory would be required in order to provide an explanation of the linguistic facts.

The type of semantic theory introduced by the transformational grammarians will be called "conceptual semantics". Conceptual semantics has roots in an intellectual tradition that is much older than transformational grammar, the "structuralism" of Ferdinand de Saussure (not to be confused with the very different "structuralism" of Bloomfield and his followers). This tradition will be discussed in the next sub-section.

In summary, by the mid-sixties the transformational grammarians felt that it was not feasible to construct an autonomous syntactic theory of a natural language. What was required instead was an overall grammar, which consisted of both a syntactic component and a semantic component. The semantic component was a "conceptual semantic" theory.

It turns out that this model of linguistic theory is in many ways compatible with Whorf's Linguistic Relativity Hypothesis. True, most of the transformationalists endorsed Chomsky's thesis of innate linguistic universals, which implies that all language users, and therefore all languages, have fundamental similarities. However, by playing down the role of universals, one can well imagine Whorf being happy with the formulation of linguistic theory that had emerged by the mid-sixties.

Unfortunately for Whorf's intellectual heritage, it turns out that there are grave problems associated with conceptual semantics. This approach will have to be abandoned, which leaves Whorf once again without the semantic theory that the Linguistic Relativity Hypothesis requires.

These themes will be expanded in the following sub-sections.

Saussurian Structuralism

The term 'structuralism' refers to two very different schools or movements within linguistics. These can be distinguished as 'American structuralism' and 'Saussurian structuralism'. The former has been discussed earlier in this chapter. The latter is in many ways antithetical to the former, especially in its views on semantic theorizing. On the other hand, Saussurian structuralism is in many ways compatible with the kind of semantic theorizing that emerged in the work of the transformational grammarians. In fact, the kind of semantic theorizing produced by the transformationalists presupposes the validity of some key principles of Saussurian structuralism. In what follows these Saussurian presuppositions will be brought out in the open.

The founding work in Saussurian structuralism is Ferdinand de Saussure's Course in General Linguistics,[99] which is actually a set of lecture notes compiled by his students and published posthumously. Saussure rejects the idea that language stands in a direct relation to reality; rather, language is related to "concepts":

> The linguistic sign unites, not a thing and a name, but a concept and a sound-image.[100]

However, it would be the gravest of mistakes, according to Saussure, to assume that there is a pre-existing set of concepts that different languages "map onto" in various ways. Without language the conceptual system is inchoate. This undifferentiated system is ordered into a set of distinct concepts only because of linguistic ordering.

---

[99] F. de Saussure, Course in General Linguistics, (1916, reprint Fontana, 1974).

[100] Ibid., 66.

> Psychologically our thought - apart from its expression in words - is only a shapeless and indistinct mass. Philosophers and linguists have always agreed in recognizing that without the help of signs we would be unable to make a clear-cut, consistent distinction between two ideas. Without language, thought is a vague uncharted nebula. There are no pre-existing ideas, and nothing is distinct before the appearance of language.
>
> Against the floating realm of thought, would sounds by themselves yield predelimited entities? No more so than ideas. Phonic substance is neither more fixed nor more rigid than thought; it is not a mold into which thought must of necessity fit but a plastic substance divided in turn into distinct parts to furnish the significance needed by thought. The linguistic fact can therefore be pictured in its totality - i.e. language - as a series of contiguous subdivisions marked off on both the indefinite plane of jumbled ideas ... and the equally vague plane of sounds...
>
> The characteristic role of language with respect to thought is not to create a material phonic means for expressing ideas but to serve as a link between thought and sound, under conditions that of necessity bring about the reciprocal delimitations of units. Thought, chaotic by nature, has to become ordered in the process of its decomposition. Neither are thoughts given material form nor are sounds transformed into mental entities; the somewhat mysterious fact is rather that "thought-sound" implies division, and that language works out its units while taking shape between two shapeless masses...
>
> Language can ... be compared to a sheet of paper: thought is the front and sound the back; one cannot cut the front without cutting the back at the same time; likewise in language, one can neither divide sound from thought or thought from sound; the division could be accomplished only abstractly, and the result would be either pure psychology or pure phonology.
>
> Linguistics then works in the borderland where the elements of sound and thought combine; their combination produces a form, not a substance.[101]

The last sentence of this quotation alludes to another major theme in Saussurian structuralism: i.e., language is composed not of a set of elements, but rather, as a set of oppositions and relations. A single phoneme or a single word cannot, according to Saussure, be characterized in isolation. Each individual phoneme or word is defined only in terms of a set of distinctions and oppositions that simultaneously define all the phonemes and all the words of the language. This means that the linguist must adopt a holistic approach to his subject matter.

---

[101] Ibid., 111-113.

> ...To consider a terms as simply the union of a certain sound with a certain concept is grossly misleading. To define it this way would isolate the term from its system; it would mean assuming that one can start from the terms and construct the system by adding them together when, on the contrary, it is from the interdependent whole that one must start and through analysis obtain its elements.[102]

For Saussure, then, language is system of internal relations. There are two amorphous realms - sound and thought - that language brings together. Language sets up a system of oppositions and relations within and between these realms such that they take on a finely grained structure. Note that Saussure is very much opposed to the Bloomfieldian principle that syntax can be done (largely) independently of semantics. Saussure insists that semantics (thought) is essential to the definition and the study of language.

Note also that Saussure is also opposed to Bloomfield's claim that semantics has something to do with the relation between language and the external world (although as we have seen, Bloomfield's semantics is somewhat ambiguous). Saussure holds that language is a link between sound and thought, rather than a relation between sound and the external world. It is useful to compare Saussure's thinking with that of Gottlob Frege. Frege held that language is a linking of sounds and "senses" (or "concepts" or "meanings"), but for Frege, these senses determine the reference - i.e., an object or set of objects in the external world - of the sounds.[103] Saussure's analogues of Frege's "senses" are "concepts" or "ideas" (although Frege would object to Saussure's psychological approach to semantics). Does Saussure subscribe to the Fregean view that concepts determine reference, thus completing the link from language to the world? Or does Saussure deny the relevance of a language-world relation altogether? The answer is not entirely clear.

---

[102] Ibid., 113.

Michael Devitt and Kim Sterelny argue that Saussurian structuralism is committed to the rejection of reference, or a language-world relation, as a key element in the understanding of language.

> The rejection of reference is central to the relational, holistic and autonomous view of language that is definitive of structuralism.[104]

Devitt and Sterelny claim that the rejection of reference is implicit rather than explicit in Saussure's own writing, but it is undeniable all the same. Saussure's comparison of language to chess[105] is indicative of this tendency. Chess can be defined solely in terms of internal relations; there is no need to consider the relation of chess pieces to the external world in describing the game. By stating that language is strongly analogous to chess, Saussure is, in effect, suggesting that the external world is equally irrelevant to the characterization of language. Devitt and Sterelny quote from a number of recent structuralists and commentators on Saussure who explicitly assert that the rejection of reference is a key doctrine of Saussurian structuralism.[106]

---

[103] G. Frege, "On Sense and Reference", Zeitschrift für Philosophie und Philosophische Kritik, 1892, 100, 25-50. Original in German, translated in Philosophical Writings of Gottlob Frege, edited by P. Geach and M. Black, (Basil Blackwell, 1977), 57.

[104] M. Devitt and K. Sterelny, Language and Reality, (The M.I.T. Press, 1987), 215.

On the other hand, many structuralist semanticists who consider themselves working within a Saussurian framework clearly do not reject the idea of a language-world relation. They argue that the task of the semantic theorist is to specify a sound-concept relation, such that the concepts will be the determiners of reference. However, the details of reference determination (the concept-world relation) are not the concern of the linguist qua linguist. Consider the following passage from Manfred Bierwisch, a structuralist semanticist:

> We will now turn briefly to the complicated question of how sentences are related, by means of their meaning, to states, processes, and objects in the universe....
>
> ...The semantic components [in a semantic theory] ...have been treated so far as purely formal elements.... It seems natural to assume that these components represent categories or principles according

---

[105] F. de Saussure, Course in General Linguistics, (1916, reprint Fontana, 1974), 88-89.

[106] M. Devitt and K. Sterelny, Language and Reality, (The M.I.T. Press, 1987), 216.

> to which real and fictitious, perceived and imagined situations and objects are structured and classified. The semantic features do not represent, however, external physical properties, but rather, the psychological conditions according to which human beings process their physical and social environments. Thus they are not symbols for physical properties and relations outside the human organism, but rather for the internal mechanisms by means of which such phenomena are perceived and conceptualized....
>
> If we interpret the semantic components in this way, their purely formal character is related to the cognitive and perceptual equipment of the human organism. This then provides the necessary interrelation of semantic structures with the surrounding universe, which is perceived and categorized according to the inherent conditions of the organism. This mediated relation between semantic structures and real situations also explains the fact that we are able to talk about things that are not present in the situation, or are purely fictitious, that we are able to form concepts corresponding to nothing in the real world[107]

Similar sentiments can be found in the work of other "Saussurian" semanticists, such as John Lyons,[108] Geoffrey Leech[109] and Ray Jackendoff.[110]

The question under consideration is the following: Does Saussurian semantics simply reject reference, as depicted in the following diagram?

---

[107] M. Bierwisch, "Semantics", in New Horizons in Linguistics, edited by J. Lyons, (Penguin, 1970), 180-182.

[108] J. Lyons, Semantics: Volume I, (Cambridge University Press, 1977), 110.

[109] G. Leech, Semantics, (Penguin, 1974), 28.

[110] R. Jackendoff, Semantics and Cognition, (The M.I.T. Press, 1985), 29-37.

Language

Sound <———> Concepts

Or, does Saussurian semantics assume that concepts are reference determiners, as depicted below?

Language

Sound <———> Concepts ———> External World

I believe that some structuralists may reject reference as charged by Devitt and Sterelny. This seems especially likely if the structuralist in question is primarily interested in literary criticism.[111] However, most linguists working in the Saussurian tradition believe that there is a language-world relation, but it is not really the business of linguistics to study it. The reference-denying structuralists that so bother Devitt and Sterelny will not be considered any further. Our focus will be on the position held by Bierwisch, Lyons, Leech and many other linguists.

As an aside, the reader may be struck that Whorf's Linguistic Relativity Hypothesis seems, on the face of it, far more compatible with Saussurian structuralist semantics than with Bloomfield's approach to semantics. I will argue that this is the case in a later section.

---

[111] For an introduction to how structuralist principles have been applied to literary criticism, see T. Eagleton, Literary Criticism, (Basil Blackwell, 1983), chapter 3 and 4.

In summary, Saussurian structuralism is compatible with a language-world relation, but it holds that natural languages in themselves are systems of internal relations. When this principle is applied to semantic theory, it means that the semantic theorist need not, and should not, try to study a relation that holds between language and the world. Rather, like the phonologist or syntactitian, the semantic theorist can study nothing other than a set of internal relations within language. The deep influence of Saussurian structuralism is apparent in the following passage:

> One of the keynotes of a modern linguistic approach to semantics is that there is no escape from language: an equation such as 'cent = hundredth of a dollar' or 'salt = NaCl' is not a matching of a linguistic sign with something outside of language; it is a correspondence between two linguistic expressions, supposedly having 'the same meaning'. The search for an explanation of linguistic phenomena in terms of what is not language is as vain as the search for an exit from a room which has no doors or windows, for the word 'explanation' itself implies a statement in a language. Our remedy, then, is to be content with exploring what we have inside the room: to study relations within language...[112]

What sort of "relations within language" must we study in order to produce a semantic theory? Saussure did not provide a systematic answer to this question,[113] although throughout the Course he did make numerous observations on the semantics of terms. However, in the relatively recent work of a large number of writers a definite answer in the Saussurian spirit has been provided. The modernized Saussurian approach is to recognize that there are a number of semantic properties that apply to both terms and sentences, and there are semantic relations that hold between pairs of terms or pairs of sentences. These semantic properties and relations are "intuitively" accessible to anyone who understands the language in question. The intuitive judgments of native speakers regarding these semantic properties and relations are considered to be the ultimate evidence for a semantic theory.

---

[112] G. Leech, Semantics, (Penguin, 1974), 5.

[113] This claim is defended by J. Culler in his introduction to F. de Saussure, Course in General Linguistics, (1916, reprint Fontana, 1974), 16.

John Lyons devotes a lengthy chapter in his Semantics to characterizing the semantic properties and relations that apply to individual lexemes (i.e. terms or words).[114] The major ones are synonymy, antonymy, semantic set membership, and hyponymy. That is, it is held that speakers of a language can recognize cases where two words such as 'little' and 'small' have the same meaning (synonymy), where two words such as 'little' and 'big' have opposing meanings (antonymy), where a set of words such as 'rose', 'tulip', 'pansy', etc. have a common element of meaning (semantic set membership), and where the meaning of one word, such as 'tulip', "contains" the meaning of another, such as 'flower' (hyponymy).

Saussurian semanticists have also identified a number of semantic properties and relations defined over sentences, rather than words. A partial list of such properties and relations is provided by Geoffrey Leech:

---

[114] J. Lyons, Semantics: Volume I, (Cambridge University Press, 1977), chapter 9.

1. X is synonymous with Y
(e.g. 'I am an orphan' is synonymous with 'I am a child and have no mother or father')
2. X entails Y
(e.g. 'I am an orphan' entails 'I have no father')
3. X is inconsistent with Y
(e.g. 'I am an orphan' is inconsistent with 'I have a father'
4. X is a tautology
(e.g. 'This orphan has no father')
5. X is a contradiction
(e.g. 'This orphan has a father')
6. X (positively) presupposes Y
(e.g. 'Is your father at home?' presupposes 'You have a father')
7. X (negatively) presupposes Y
(e.g. 'If he had a father, things would be different' negatively presupposes 'He has a father')
8. X is semantically anomalous
(e.g. 'The orphan's father drinks heavily')[115]

Leech calls these types of statements the "basic statements" on which semantic theorists must base their theories. The idea is that semantic theorists assemble a set of such basic statements that seem intuitively correct to themselves and to other speakers of the language. To this must be added the intuitive judgements that speakers make about the semantic properties of individual words. Then by using certain fairly obvious inferences from this data (which I will not go into), various semantic conclusions can be drawn and woven together into an overall semantic theory for the language in question. For example, the data given in Leech's example supposedly allows us to conclude that 'orphan' means 'child without a father or mother'.

---

[115] G. Leech, Semantics, (Penguin, 1974), 85.

To conclude, Saussurian structuralism has many themes, but one of the central ones is that language is a relation between sound and a conceptual realm. Semantics is the study of the detailed sound-concept relation in particular languages, and the method for this study is to focus on certain internal relations between units in the language. In recent times this Saussurian methodological principle has evolved into a focus on semantic properties and relations that are defined over the set of terms and sentences of the language. As we shall see, this Saussurian approach to semantics was incorporated into the work of transformational grammarians in the mid-sixties.

Componential Analysis

We saw in the last sub-section that hyponymy is one of the semantic relations that the Saussurian semanticist uses as basic evidence in the construction of a semantic theory. Hyponymy, to recall, is a relation that holds between two lexemes when the meaning of one is contained within the meaning of another. For a hyponymous pair like 'tulip' and 'flower', it has become customary to call the more specific term, like 'tulip', a "hyponym", while the more general term, like 'flower', is called the "subordinate term".[116]

Reflection on the phenomenon of hyponymy leads to the conjecture that there are elements or components of meaning that reappear in many different lexemes. The meaning of 'flower' appears not only in the meaning of 'tulip', but also in the meaning of 'rose', pansy', etc. Furthermore, all these lexemes arguably contain the meaning of 'organic', which is also contained in lexemes like 'muskrat' and 'earthworm'. The extent of this phenomenon is considered something that should not go unnoticed in semantic theory, and this consideration has led to a theoretical approach known as "componential analysis". As we have just seen, componential analysis is almost a consequence of the Saussurian approach to semantics, and therefore:

---

[116] Ibid., 101.

> it is probably true to say that the majority of [Saussurian] structural semanticists subscribe nowadays to some version or other of componential analysis. This approach to the description of the meaning of words and phrases rests upon the thesis that the sense of every lexeme can be analyzed in terms of a set of more general sense-components (or semantic features), some or all of which will be common to several different lexemes in the vocabulary. In so far as componential analysis is associated with conceptualism, the sense-components (for which there is so far no generally accepted term) may be thought of as atomic, and the sense of particular lexemes as molecular, concepts.[117]

An example will help make this clear. Suppose we are trying to account for the semantics of the following terms in English:

'man_1', as in 'Modern man evolved fairly recently'
'man_2', as in 'That man insulted her'
'woman'
'female'
'boy'
'girl'
'adult'
'child'

A componential analysis would hold that the meanings of these terms can be accounted for in terms of three atomic meanings, or "semantic features", symbolized by the capitalized words HUMAN, ADULT and MALE. Each of these features represents a binary opposition of meanings; that is, +MALE represents the meaning of 'male', whereas -MALE represents the meaning of 'female'. The meanings of this list of English words can then be specified as follows:[118]

---

[117] J. Lyons, Semantics: Volume I, (Cambridge University Press, 1977), 317.

'man$_1$'	+HUMAN		
'man$_2$'	+HUMAN	+ADULT	+MALE
'woman'	+HUMAN	+ADULT	-MALE
'female'			-MALE
'boy'	+HUMAN	-ADULT	+MALE
'girl'	+HUMAN	-ADULT	+MALE
'adult'	+HUMAN	+ADULT	
'child'	+HUMAN	-ADULT	

Componential analysis allows a formalization of the concept of hyponymy. A pair of terms are hyponymous if the componential analysis of the subordinate term is a subset of the componential analysis of the hyponym. Thus, 'man$_1$' is a subordinate term to 'adult', since the componential analysis of 'man$_1$' is a subset of the componential analysis of 'adult'.

There are a number of problematic issues that must be addressed by the advocate of componential analysis. One of the central issues is the degree of atomicity that the componential theorist should engage in. In other words, when should the componentialist stop breaking concepts up into more basic concepts? Leech's answer is that:

> a useful rule of thumb is to recognize an opposition of meaning [i.e., a binary semantic feature such as HUMAN] whenever it proves its value by allowing us to make generalizations covering a range of lexical items.[119]

---

[118] G. Leech, Semantics, (Penguin, 1974), 96-97.

[119] Ibid., 99.

However, the consequence of this approach is that componential analyses will not be able to give comprehensive accounts of the meaning of lexemes, since in any language it is a contingent possibility that a term may involve a "concept" that does not appear in any other term of the language. If we use Leech's criterion for determining the limits of atomization, we will never be able to capture these unique concepts in a componential analysis. Some theorists who accept Leech's criterion have gone on to say that componential analysis must be supplemented with other theoretical machinery in order to specify the meanings of terms. For example, in their classic article "The Structure of a Semantic Theory", Katz and Fodor state that the meaning of a term must, in general, be specified by identifying both "semantic markers", which are the "components" of a componential analysis, and "distinguishers", which represent "concepts" that appear only in single terms.

> The distinction between markers and distinguishers is meant to coincide with the distinction between that part of a lexical item which is systematic for the language, and that part which is not. In order to describe the systematicity in the meaning of a lexical item, it is necessary to have theoretical constructs whose formal interrelations compactly represent this systematicity. The semantic markers are such constructs. The distinguishers, on the other hand, do not enter into theoretical relations within a semantic theory. The part of the meaning of a lexical item that a dictionary represents by a distinguisher is the part of which a semantic theory offers no general account.[120]

---

[120] J. J. Katz and J. A. Fodor, "The Structure of a Semantic Theory", Language, 1963, 39, 170-210. Reprinted in Readings in the Psychology of Language, edited by L. A. Jakobovits and M. S. Miron, (Prentice-Hall, 1967), 413-414.

Others have argued against this dualism between markers and distinguishers, holding that componential analysis should atomize meaning to the point where very term in the language can be given a componential analysis that exhausts its meaning. Thus, Anna Wierzbicka advocates a "minimalization principle" in componential analysis according to which

> the list of indefinables [i.e. components] must be as small as possible; it should contain only those elements which are really absolutely essential while being at the same time adequate to explain all utterances.[121]

This debate obviously cannot be decided on the basis of pre-theoretical intuitions about language. The decision to adopt one course or another must be motivated by theoretical considerations, such as simplicity and scope of explanatory power. Fortunately, we need not resolve this debate at this juncture. The significant point is simply that Saussurian semantics calls attention to the fact that the meanings of terms may be "contained" in the meanings of other terms. Componential analysis was invented as a means of capturing these relations and had been adopted by almost all Saussurian semanticists. Whether componential analysis must be supplemented with "distinguishers" in order to specify word meaning is a relatively minor point.

Semantics in Generative Grammar

Although the Saussurian structuralist approach to semantics has been practised and elaborated continuously since Saussure's death, it was held in low repute by most American linguists during the three decade span of Bloomfield's influence. (Whorf, consequently, did not make use of Saussurian terminology.)

---

[121] A. Wierzbicka, Semantic Primitives, (Athenaum Verlag, 1972), 13.

However, as we have seen, Chomsky's critique of the Bloomfieldian approach to syntax resulted in a willingness of linguists to once again engage in theory construction in linguistics. This soon spilled over into semantics, and in fact, the Saussarian approach to semantics was assimilated into the Chomskyan framework of transformational grammar in a classic article written by J. J. Katz and J. A. Fodor in 1963.[122]

Katz and Fodor's Theory

Katz and Fodor justified their excursion into semantics by invoking the Chomskyan principle that a linguistic theory should account for a speaker's competence. That is, the theory should be a formalization of the idealized knowledge of the language that a competent speaker possesses. One type of competence that an idealized speaker possesses is the ability to resolve ambiguous sentences. An example of an ambiguous sentence is:

(i) Flying planes can be dangerous

which can be disambiguated by the following two paraphrases:

(ii) Flying planes is sometimes dangerous

(iii) Flying planes are sometimes dangerous.

Chomsky's purely syntactic theory in Syntactic Structures was capable of accounting for this ambiguity by proposing that there are two distinct syntactic derivations of (i), and the structural differences between these derivations account for the two readings.

---

[122] J. J. Katz and J. A. Fodor, "The Structure of a Semantic Theory", Language, 1963, 39, 170-210. Reprinted in Readings in the Psychology of Language, edited by L. A. Jakobovits and M. S. Miron, (Prentice-Hall, 1967).

However, Katz and Fodor observed that some ambiguous sentences cannot be demonstrated to have structural ambiguities. For example:

(iv) The bill is large

is ambiguous, but only because the lexical item 'bill' is ambiguous. Chomsky's theory in Syntactic Structures is not capable of explaining this sort of ambiguity, and therefore fails as a complete linguistic description of English if a criterion of adequacy on linguistic descriptions is that they account for a native speaker's competence.

Chomsky's purely syntactic theory must therefore be supplemented, argued Katz and Fodor. They proposed the following maxim: "Linguistic description minus grammar [=syntax] equals semantics". That is, they held that by supplementing Chomskyan syntax with a semantic theory they would be able to realize the Chomskyan objective of accounting for the idealized knowledge possessed by a speaker of a language.

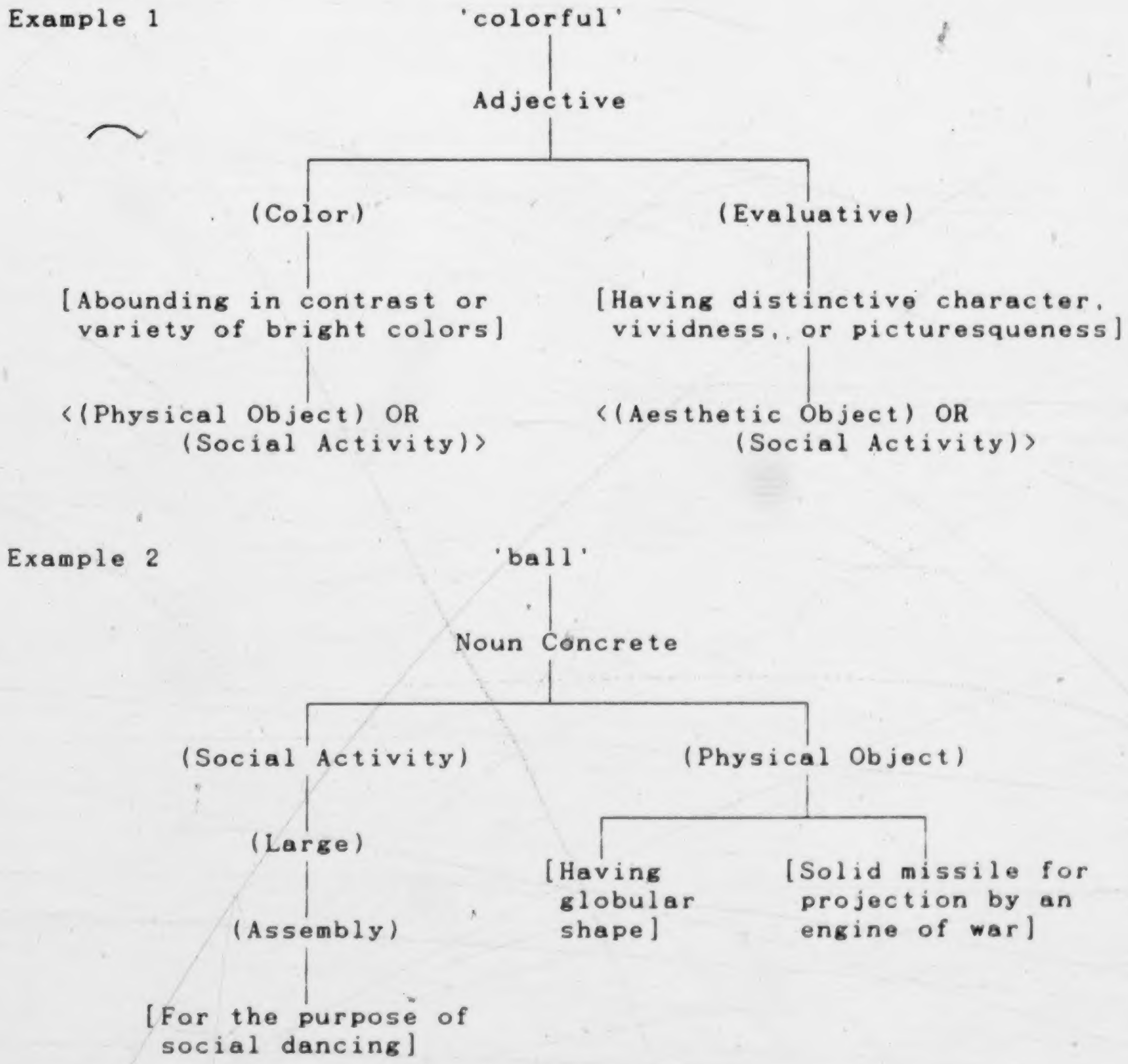
The output of a transformational sentence-syntax will be an infinite set of sentences, each of which is associated with a phrase structure.[123] If a sentence is syntactically n-ways ambiguous, then it will be output n times from the sentence-syntax, each time paired with a distinct structural description. Katz and Fodor argued that these structurally described sentences should be regarded as input to the semantic component of a complete generative grammar (which also contains the transformational sentence-syntax, of course, and a phonological component). The output of the semantic component will be a "semantic representation" of each structurally distinct sentence. If an input sentence is semantically ambiguous, as in example (iv) above, then the semantic component will generate m semantic representations; one for each of the m ways in which the input sentence is semantically ambiguous. The semantic interpretation of a sentence S will be the conjunction of the semantic interpretations of each of the syntactic derivations of S, keeping in mind that each derivation may have more than one semantic interpretation.

---

[123] Even if a sentence was generated in part by the application of non-phrase structure transformational rules, it will still have, at each stage of its derivation, a unique phrase structure provided that the transformational rules have been appropriately formulated.

The semantic component consists of two major components: the dictionary component and the projection rule component. The dictionary contains an entry for each lexical item in the language. Two examples of dictionary entries are given below:

Figure 2.3.1
Katz and Fodor's Dictionary Entries


Example 1
'colorful'
Adjective
(Color)
(Evaluative)
[Abounding in contrast or variety of bright colors]
[Having distinctive character, vividness, or picturesqueness]
<(Physical Object) OR (Social Activity)>
<(Aesthetic Object) OR (Social Activity)>
Example 2
'ball'
Noun Concrete
(Social Activity)
(Physical Object)
(Large)
(Assembly)
[For the purpose of social dancing]
[Having globular shape]
[Solid missile for projection by an engine of war]

Each path in a Katz and Fodor dictionary entry represents a distinct meaning of the lexical item. 'Colorful' has two meanings, and 'ball' has three, according to the entries given in figure 2.3.1.

Each dictionary entry consists of the lexical item, followed by the syntactic categories, represented by "grammatical markers", to which that item belongs. The examples given above belong to only one syntactic category, but the English word 'butter' is both a noun and a verb, and consequently its Katz and Fodor dictionary entry would contain two branches. Note also that a path can contain more that one grammatical marker if necessary.

Next, all the semantic markers associated with that item are entered on the path. Semantic markers, which are enclosed in parentheses, are the "components" of componential analysis. That is, they are intended to represent atomic concepts that may also appear in the dictionary entries of other lexical items. As can be seen from the leftmost path of the 'ball' example, a path may contain several semantic markers.

Next, the distinguisher is included on the path. These are enclosed in square brackets. As mentioned in the previous subsection, a Katz and Fodor distinguisher represents the conceptual content of the lexical item that is unique to it, and which does not reappear in any other lexical item, and therefore is not represented by any semantic marker.

Finally, the dictionary entry will contain information that restricts the semantic contexts into which that lexical item can be introduced. This information is enclosed in angle brackets, and is always represented as a Boolean relation defined over semantic markers. The role of this selection restriction information will become clearer when we examine the projection rule component, to which we now turn.

The purpose of the dictionary is to provide a formalization of the meaning of lexical items. It must represent every sense that the lexical item can have in any sentence of the language. The purpose of the projection rules, on the other hand, is to determine the meaning of sentences as a function of the meanings of the lexical items that they contain. The projection rule component consists of a finite number of rules that generate what Katz and Fodor call an "amalgamated path", which is a representation of sentence meaning. Note, by the way, that this represents an advance over the componential analysis up that time. Prior to Katz and Fodor's paper, most componential analysts were content to restrict their semantic analyses to lexical items. Katz and Fodor realized that sentence-meaning must be explained in any complete semantic theory. No doubt their exposure to Chomsky's approach to syntax led them to pursue this objective in semantics.

The details of the projection rule component would require a lengthy explanation that is not required in the present context. It is sufficient to point out three features of the projection rules:

(1) There is one projection rule for every rule of phrase structure composition that is included in the sentence-syntax. For example, assuming that the syntax allows the concatenation of 'colorful' and 'ball' into the noun phrase 'colorful ball', then

there will be a corresponding projection rule that allows us to "amalgamate" the paths (i.e., dictionary entries) of 'colorful' and 'ball' into an "amalgamated path" that represents the meaning of 'colorful ball'.

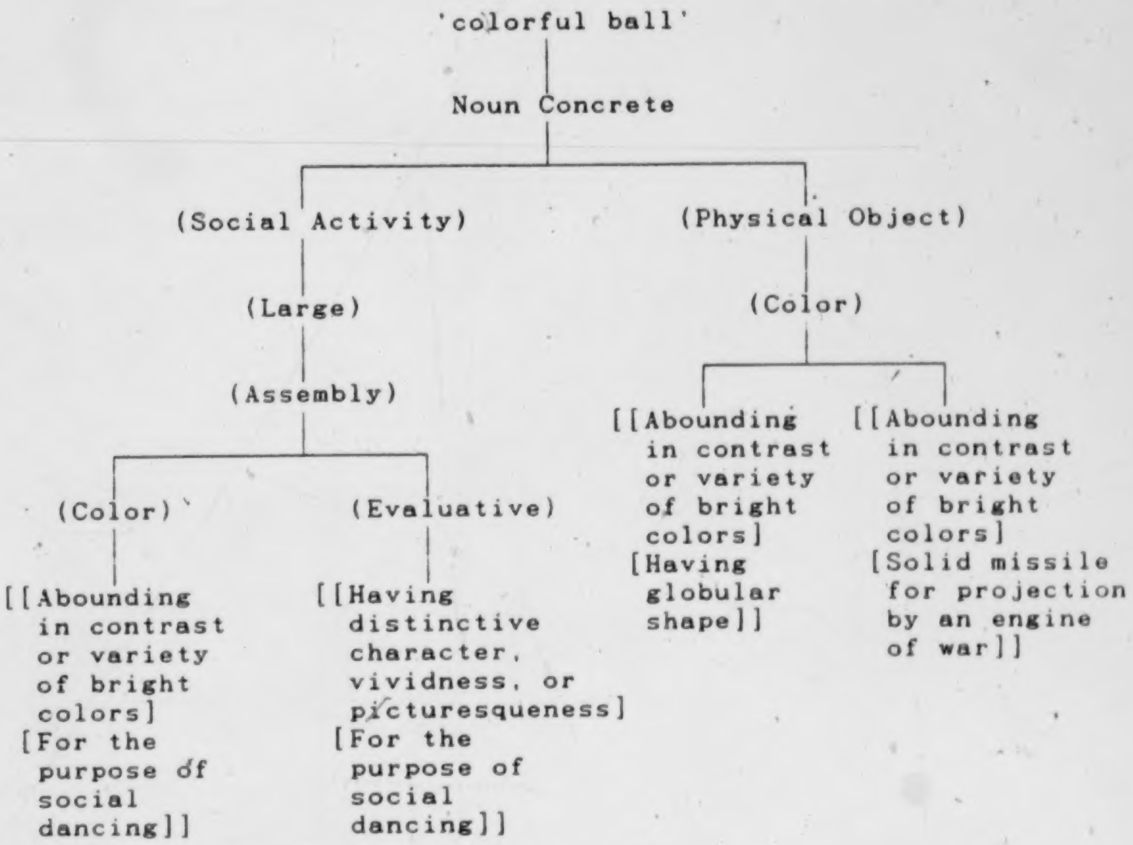
(2) The projection rules function as "filters". An example will make this clear. We saw in figure 2.3.1 that Katz and Fodor associate two meanings with 'colorful' and three with 'ball'. Therefore there are six possible "amalgamated paths" that can be associated with 'colorful ball'. However, the projection rule that Katz and Fodor has formulated for forming these sorts of noun phrases will filter out two of these amalgamated paths. Specifically, their rule does not allow for the second meaning (right-hand path) of 'colorful' to be amalgamated with the first two meanings (left-hand and middle paths) of 'ball'.[124]

(3) Finally, the Katz and Fodor projection rules assign the same sort of "meanings" to syntactically complex units as they do to syntactically simple units, i.e., lexical items. For example, Katz and Fodor's projection rule for attributive noun phrases assigns four meanings to the phrase 'colorful ball' as represented in the following diagram:

---

[124] J. J. Katz and J. A. Fodor, "The Structure of a Semantic Theory", Language, 1963, 39, 170-210. Reprinted in Readings in the Psychology of Language, edited by L. A. Jakobovits and M. S. Miron, (Prentice-Hall, 1967), 422.

Figure 2.3.2
Katz and Fodor's Representation of the Meaning of a Syntactically Complex Unit


'colorful ball'
Noun Concrete
(Social Activity)
(Physical Object)
(Large)
(Color)
(Assembly)
[[Abounding in contrast or variety of bright colors] [Having globular shape]]
[[Abounding in contrast or variety of bright colors] [Solid missile for projection by an engine of war]]
(Color)
(Evaluative)
[[Abounding in contrast or variety of bright colors] [For the purpose of social dancing]]
[[Having distinctive character, vividness, or picturesqueness] [For the purpose of social dancing]]

With the exception of the angle-bracketed selection restrictions (which were used to filter out the two other possible paths) these paths have the same structure as a dictionary entry for a single lexical item. Similarly, verb phrases and even sentences will have their meanings represented by similar amalgamated paths. For example, one of the meanings that Katz and Fodor's projection rules provide for the sentence 'The man hits the colorful ball' is given by the following path:

> 'The man hits the colorful ball' -> Sentence -> [Some contextually definite] -> (Physical Object) -> (Human) -> (Adult) -> (Male) -> (Action) -> (Instancy) -> (Intensity) -> [Strikes with a blow or missile] -> [Some contextually definite] -> (Physical Object) -> (Color) -> [[Abounding in contrast or variety of bright colors][Having globular shape]][125]

The Impact of Katz and Fodor's Theory

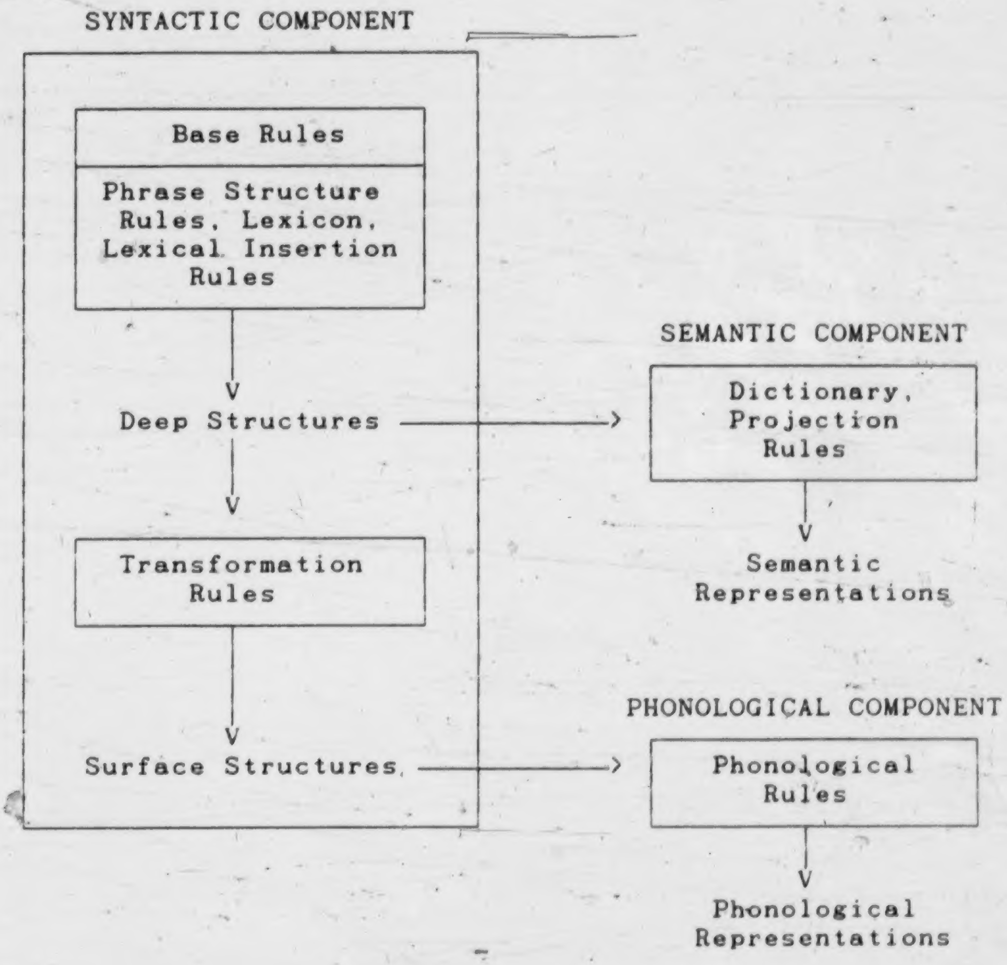
Katz and Fodor's paper was enormously influential. It effected the synthesis of the Saussurian approach to semantics with the Chomskyan approach to syntax. Chomsky himself incorporated a semantic component in his next major work Aspects of the Theory of Syntax.[126] Chomsky proposed the following overall structure for a generative grammar:

---

[125] Ibid., 427. Katz and Fodor's projection rules generate three other meanings or "amalgamated paths" for this sentence.

[126] N. Chomsky, Aspect of the Theory of Syntax, (The M.I.T. Press), 1965, 15-18.

Figure 2.3.3
Chomsky's Organization of a Generative Grammar (1965)


SYNTACTIC COMPONENT
Base Rules
Phrase Structure Rules, Lexicon, Lexical Insertion Rules
Deep Structures
Transformation Rules
Surface Structures
SEMANTIC COMPONENT
Dictionary, Projection Rules
Semantic Representations
PHONOLOGICAL COMPONENT
Phonological Rules
Phonological Representations

That is, Chomsky argued in Aspects that transformations preserve meaning (which required some revision of the transformational rules that were specified in Syntactic Structures), and therefore the output of the "base" component of the syntax (i.e., everything but the transformational rules) would be the input to the semantic component. Note that the entire grammar produces a pairing of sounds with meanings – a Saussurian objective for linguistic theory.

The exact relation between syntax and semantics became a topic of great debate within linguistics over the next ten to fifteen years. Chomsky revised his theory so that surface structures as well as deep structures could be input into the semantic component. This came to be known as the Extended Standard Theory. Another group of theorists, the "generative semanticists", argued that the base component of a grammar directly generates semantic representations, and that deep structures as a distinct syntactic representation were not required. These theorists also altered the notion of a semantic representation, arguing that it is a tree-like structure with complexes of semantic markers as terminal elements. But while there were many hotly debated differences between and even within these schools, and while the details of syntactic and semantic representation were continuously modified, the overall conception of semantics within the generative framework was clearly established by Katz and Fodor. What Katz and Fodor did was to extend the Saussurian "conceptual" approach to semantics from individual words to whole sentences. The debates that they kicked off for the next two decades were essentially on matters of detail.[127]

---

[127] Recently some of Chomsky's followers have renounced semantics altogether. See, for example, N. Hornstein, Logic as Grammar, (The M.I.T. Press, 1984), chapters 6 and 7. However, in spite of these trends, the Saussurian/Chomskyan synthesis is very

Why Katz and Fodor's Semantics is Inadequate

Saussurian structuralism proposed that the meanings of individual words should be represented as complexes of concepts. Katz and Fodor showed how this sort of analysis could be extended to phrases and whole sentences, so that the meaning of these larger units can also be represented as (slightly more complex) complexes of concepts.

However, it has been argued by several commentators that this approach to semantics amounts to nothing more than translation into a canonical language. David Lewis, for example, writes:

> Semantic markers are *symbols*: items in the vocabulary of an artificial language we may call *Semantic Markerese*. Semantic interpretation by means of them amounts merely to a translation algorithm from the object language to the auxiliary language Markerese. But we can know the Markerese translation of an English sentence without knowing the first thing about the meaning of the English sentence: namely, the conditions under which it would be true. Semantics with no treatment of truth conditions is not semantics. Translation into Markerese is at best a substitute for real semantics, relying either on our tacit competence (at some future date) as speakers of Markerese or on our ability to do real semantics at least for the one language Markerese. Translation into Latin might serve as well, except insofar as the designers of Markerese may choose to build into in useful features - freedom from ambiguity, grammar based on symbolic logic - that might make it easier to do real semantics for Markerese than for Latin.[128]

---

much alive as will be shown in the discussion of Ray Jackendoff'x theory a few pages hence.

[128] D. Lewis, "General Semantics", in *Semantics of Natural Language*, 2nd ed., ed. by D. Davidson and G. Harman, (D. Reidel Publishing, 1972), 169-170. For similar criticisms see G. Evans and J. McDowell, "Introduction", in *Truth and Meaning*, ed. by G. Evans and J. McDowell, (Oxford University Press, 1976), vii-xii, and H. Putnam, "Is Semantics Possible?", in his *Mind, Language and Reality*, (Cambridge University Press, 1975), 146.

This passage makes two points: (1) that the so-called semantic program practised by generative grammarians amounts to no more than the specification of a translation algorithm, and (2) that a translation algorithm is not a semantic theory. We have already considered, and endorsed, point (2) in our discussion of Quine's semantic program. The issue that must now be considered is whether Lewis' first point is correct.

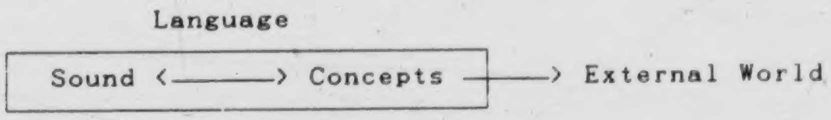
There is no doubt that taken by themselves and without further elaboration, the theory of Katz and Fodor and the later Extended Standard Theory and the theories of the generative semanticists did nothing more than define a relation between phonological representations of sentences and their semantic representations. Given that the semantic representations are nothing more than complexes of symbols in a standard "Markerese" vocabulary, then it is indeed true that when taken by themselves and without further elaboration, these theories are nothing more than translation algorithms.

However, the possibility remains open that the markerese representations may, when placed in a larger theoretical context, have true semantic properties. This possibility may be exploited by claiming that semantic markers (and structures built out of them) are not merely formal symbols, but are representations of concepts (and conceptual structures), and that these concepts can do what mere translation cannot; that is, they can perform the real semantic function of establishing a relation between language and the world.

In fact, many practitioners of the semantic markerese approach have assumed exactly that. In her book on semantics in generative grammar, Janet Dean Fodor writes:

> ... the semantic component of a generative grammar ... assigns CONCEPTS to expressions as the means of specifying their meanings.[129]
>
> ... a semantic marker is supposed to designate a concept.[130]

Many other advocates of the semantic marker approach have provided similar theoretical context for the markerese approach.[131] The question that must now be considered is: Is this theoretical context defensible? Can we make sense of the notion that a Katz and Fodor style semantics theory establishes a relation between sounds and conceptual structures, and that these conceptual structures are related to the world in some way? In other words, can we make sense of the following diagram (which we previously used to represent the Saussurian semantic program)?



Katz and Fodor's theory and the theories that followed it are inadequate insofar as they do not explain the concept-world relation. The next subsection examines a theory within the tradition of Saussurian semantics that attempts to provide an explanation of this relation.

---

[129] J. D. Fodor, Semantics: Theories of Meaning in Generative Grammar, (Harvard University Press, 1977), 17.

[130] Ibid., 174.

[131] For a book length treatment see R. L. Peterson, Concepts and Language, (Mouton, 1973).

Jackendoff's Semantic Theory

Ray Jackendoff has developed a semantic theory that conforms with the basic principles of Katz and Fodor's synthesis of Saussurian structural semantics and generative grammar, but Jackendoff has gone beyond Katz and Fodor and has tried to provide a psychological explanation of the notion of "concept" and the concept-world relation.[132]

Following the lead of Chomsky, Jackendoff holds that linguistic theories should be an explication of the competence of an ideal speaker. The semantic component of a linguistic theory, then, should be considered as an explication of the idealized speaker's capability to make judgements of synonymy, hyponymy, and so on. Jackendoff extends this point, and argues that human semantic competence is not exhausted by these purely linguistic judgements. Semantic competence is closely linked to other cognitive abilities. Introducing the book in which he develops this theory, Jackendoff writes:

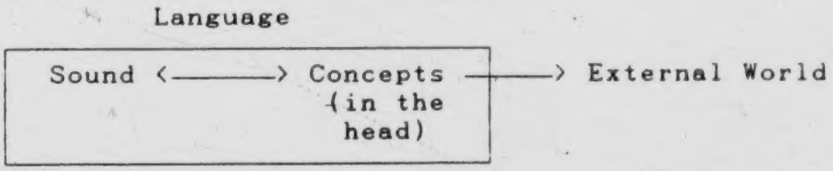
> This book is intended to be read from two complementary perspectives. From the point of view of linguistics and linguistic philosophy, the question is:

---

[132] R. Jackendoff, _Semantics and Cognition_, (The M.I.T. Press, 1983). Note that this 1983 theory is significantly different from his earlier work, _Semantic Interpretation in Generative Grammar_, (The M.I.T. Press, 1972). The 1972 book is frequently cited in the literature, so the reader is cautioned not to confuse the two theories. A summary of the 1983 theory can be found in R. Jackendoff, "Conceptual Semantics", in U. Eco, M. Santambrigio, and P. Violi, eds., _Meaning and Mental Representations_, (Indiana University Press, 1988).

> What is the nature of meaning in human language, such that we can talk about what we perceive and what we do? From the point of view of psychology, the question is: What does the grammatical structure of natural language reveal about the nature of perception and cognition?
>
> My thesis is that these two questions are inseparable: to study semantics of natural language is to study cognitive psychology. I will show that, viewed properly, the grammatical structure of natural language offers an important new source of evidence for the theory of cognition.[133]

More precisely, Jackendoff's thesis is that the semantic representations of a Katz and Fodor style semantic theory are actually cognitive structures that are used by speakers to perceive and conceive the world. To put it bluntly, concepts are in people's heads. Our knowledge of language consists in our ability to map sounds to these concepts. These very same concepts are then used to categorize the things and events in the world. The following diagram, according to Jackendoff, should be interpreted psychologically.



David Lewis rejected Katz and Fodor's theory on the grounds that it amounts to no more than a translation manual. However, it is not possible to write off Jackendoff's theory so quickly. Jackendoff's position is that a complete cognitive psychology will include not only a psycholinguistic theory of how sound is mapped to concepts, but also a more general theory of how organisms can internally represent the world with concepts. By combining the psycholinguistic theory with the general theory of concept application, it is possible to construct a mapping between sounds and the world, which is a necessary requirement on semantic theories, according to Lewis.

---

[133] R. Jackendoff, Semantics and Cognition, (The M.I.T. Press, 1983), 3.

A number of aspects of Jackendoff's theory will now be examined. Although Jackendoff is not the only theorist who has elaborated the psychological approach just described, his work is especially appropriate because of its remarkable affinity with Whorf's views.

The Conceptual Structure Hypothesis

Jackendoff holds that human cognitive abilities are mediated by "conceptual structures" which have representational properties. He makes the following empirical claim about these structures:

> The Conceptual Structure Hypothesis
>
> There is a single level of mental representation, conceptual structure, at which linguistic, sensory, and motor information are compatible.[134]

This is an empirical claim, according to Jackendoff, since it is possible that each cognitive modality (sight, hearing, motor control, etc.) could have its own separate system of representation, with inter-modal interfaces as appropriate.

Jackendoff goes on to say:

> Conceptual Structure must be rich enough in expressive power to deal with all things expressible by language. It must also be rich enough in expressive

---

[134] Ibid., 17.

> power to deal with the nature of all the other modalities of experience as well - no simple matter. In order to give some formal shape to the problem, I will assume that the possible conceptual structures attainable by a human being are characterized by a finite set of conceptual well-formedness rules. I will further assume that these rules are universal and innate - that everyone has essentially the same capacity to develop concepts - but that the concepts one actually develops must depend to some extent on experience.[135]

Jackendoff believes in the existence of universal and innate conceptual well-formedness rules because, in his opinion, it would be impossible to explain concept learning without the postulation of these rules. Jackendoff fully endorses the following argument by Jerry Fodor on the nature of the acquisition of natural language:

> Learning a language, (including, of course, a first language) involves learning what the predicates of the language mean. Learning what the predicates of a language mean involves learning a determination of the extension of these predicates. Learning a determination of the extension of the predicates involves learning that they fall under certain rules (i.e., truth rules). But one cannot learn that P falls under R unless one has a language in which P and R can be represented. So one cannot learn a language unless one has a language. In particular, one cannot learn a first language unless one already has a system capable of representing the predicates in that language and their extensions. But, on pain of circularity, that system cannot be the language that is being learned. But first languages are learned. Hence, at least some cognitive operations are carried out in languages other than natural languages.[136]

Fodor's well-known (and controversial) conclusion is that there must be an innate language of thought in which language learning is conducted. Jackendoff suggests that Fodor's argument can be generalized to apply to all conceptual learning, even concepts that are not expressed linguistically.

---

[135] Ibid., 17.

[136] J. A. Fodor, The Language of Thought, (Thomas Y. Crowell, 1975), 63-64.

There is a significant difference between Fodor and Jackendoff that must be pointed out. Fodor argues that the innate language of thought must be characterized by a very large number of primitive predicates that are very rich in expressive power.[137] This position implies that the conceptual resources of natural languages cannot be very different from one another, since they are all based on the underlying innate language of thought - a very anti-Whorfian consequence. Jackendoff, on the other hand, holds that what is innate is a set of conceptual well-formedness rules. As indicated in the last passage quoted from Jackendoff, this means that if individuals have different experiences, then the rules will operate on those experiences to produce different concepts in those individuals. Jackendoff includes experiences with natural language among those that may affect conceptual development:

> The grammatical and lexical choices of one's native language quite possibly help shape the relative salience of concepts one develops, so there is room in the theory of a certain amount of 'Whorfian' variation in concepts due to linguistic experience.[138]

In summary, Jackendoff proposes, as an empirical hypothesis, that there is an innate system of conceptual well-formedness rules that interact with experience to cause the development of a system of conceptual structures which act as a common representational system for linguistic, sensory and motor systems.

---

[137] Ibid., 79-82.

[138] R. Jackendoff, Semantics and Cognition, (The M.I.T. Press, 1983), 242.

Semantic Structure is Conceptual Structure

The Katz and Fodor approach to semantic theory proposes that each sentence in a natural language be given a semantic representation. This semantic representation can, as David Lewis has pointed out, be viewed as a translation into a special language called Markerese. Jackendoff points out that there are two ways in which the semantic representations of linguistic entities could be related to the common system of conceptual structures. On the one hand, semantic representations could somehow be mapped into the conceptual system. On the other hand, semantic representations might simply be identified with the conceptual system.

Jackendoff endorses the latter option. His argument for identifying semantic representation is quite simple. He first argues that non-linguistic categorization (for example, the judgment that an object is a dog, say) entails certain things about the structure and operation of the conceptual system.[139] He then argues that the semantic representation of a great variety of sentences (in English, where all his examples come from) requires the postulation of a system with exactly the same structure and operation.[140] Finally, he argues that most of the semantic properties, e.g., synonymy, entailment, inconsistency, etc., that, according to tradition, a semantic theory must explain, can, in fact, be explained by his theory of conceptual structures.[141]

---

[139] Ibid., chapter 5.

[140] Ibid., 95-103.

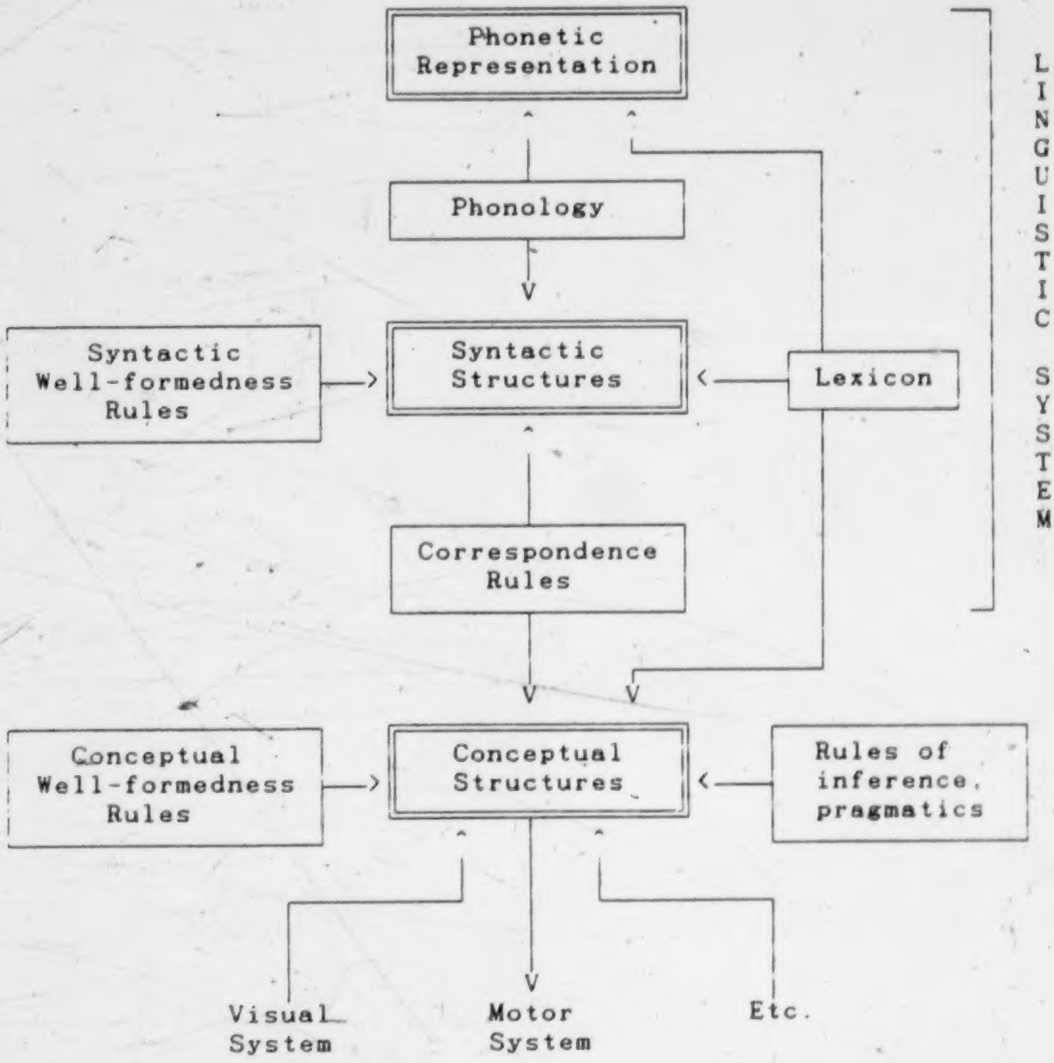
This argument is not intended to be decisive; rather, it is intended to provide some motivation for the identification of semantic representation with conceptual structure, an identification that Jackendoff believes will be further vindicated by the overall adequacy of the theory.

Given that semantic representations have been identified with conceptual structures, a natural question to ask is: what is the relation between syntactic structures and semantic representations (= conceptual structures). It will be recalled that Katz and Fodor postulated "projection rules" that linked syntactic structures to semantic representations, and rules of this nature were incorporated into Chomsky's 1965 theory. Jackendoff similarly postulates a set of rules that link syntax and semantics, except that he calls them "correspondence rules". These rules, along with other components of Jackendoff's theory, are depicted in the following diagram:

---

[141] Ibid., 104-105.

Figure 2.3.4
Jackendoff's Organization
of Linguistic and Cognitive Structures[142]


Phonetic Representation
Phonology
Syntactic Well-formedness Rules
Syntactic Structures
Lexicon
Correspondence Rules
Conceptual Well-formedness Rules
Conceptual Structures
Rules of inference, pragmatics
Visual System
Motor System
Etc.
LINGUISTIC SYSTEM

Jackendoff argues that the correspondence rules do very little work. That is, syntactic and semantic structures are structurally very similar, and consequently the correspondence rules perform a rather straightforward mapping.

---

[142] Adapted from Ibid., 21.

The argument for this close relation between syntax and conceptual structure is based on Chomsky's claims about language acquisition; viz., that there is a huge difference between the "poverty of the stimulus" available to the child and what he finally learns, and that human mental structures must therefore be constituted in such a way that the language learner can bridge this gap. Consequently, it is reasonable to assume, according to Jackendoff, that we are psychologically constituted with a close mapping between conceptual structures and grammatical structures so that language learning will be made easier.[143] In fact, Jackendoff elevates this argument into a methodological principle:

> The Grammatical Constraint [a constraint on semantic theories] says that one should prefer a semantic theory that explains otherwise arbitrary generalizations about the syntax and the lexicon.[144]

Jackendoff, a linguist by trade, advocates a version of what is known as the Extended Standard Theory of sentence-syntax.[145] This theory includes a theory of syntactic categories, known as the X-Bar Theory. According to the X-Bar Theory:

> a primary distinction is customarily made between the lexical categories (or parts of speech) - e.g., Noun (N), Verb (V), Adjective (A), and Preposition (P)

---

[143] Jackendoff takes this argument from Fodor's The Language of Thought. See R. Jackendoff, *Semantics and Cognition*, (The M.I.T. Press, 1983), 13-14.

[144] Ibid., 13.

[145] R. Jackendoff, *X-Bar Syntax: A Study of Phrase Structure*, (The M.I.T. Press, 1977).

> - and phrasal categories - e.g., Noun Phrase (NP), Verb Phrase (VP), Adjective Phrase (AP), Prepositional Phrase (PP), and Sentence (S). Each phrasal category contains a head - a member of one of the lexical categories, plus a variety of possible modifiers, which are typically other phrasal categories. Corresponding to each lexical category there is a major phrasal category, that phrasal category which maximizes the possible modifiers of the lexical category. The major phrasal category corresponding to N is NP; that corresponding to V is S.[146]

One way of looking at this syntactic theory is that a head is a "function" that takes "arguments" (modifiers) to yield a "value" (an instance of a major phrasal category). The phrase structure of a typical sentence can therefore be viewed as a hierarchy of such function-argument structures.

According to Jackendoff, the function-argument structure of the X-Bar Theory of syntactic categories is mirrored exactly in conceptual structure. The lexical head of a major phrasal category in a sentence will correspond to a conceptual function that takes the conceptual analogue of the syntactic modifiers to yield a conceptual analogue of a major phrasal category. The semantic representation of a sentence, according to Jackendoff, will therefore be a hierarchy of conceptual function-argument structures that is isomorphic to the syntactic function-argument structures postulated by the X-Bar Theory.[147] Another way of putting this is that the conceptual well-formedness rules are isomorphic to the syntactic well-formedness rules of the X-Bar Theory.[148]

---

[146] R. Jackendoff, Semantics and Cognition, (The M.I.T. Press, 1983), 63-64.

[147] Ibid., 67.

The question that naturally arises is how do these conceptual structures relate to the world, and how are the expressions of natural language thereby related to the world? In order to answer this question we have to first consider another aspect of Jackendoff's theory.

Jackendoff's Kantianism

The most straightforward way of explaining the language-world relation in the context of Jackendoff's theory would appear to be the following: The expressions of a natural language are mapped to conceptual structures in the heads of the speakers of the language by means of correspondence rules, which are also in their heads. The conceptual structures are reference determiners; that is, people pick out objects, events, sets, etc., in the world by determining what concepts they fall under. A linguistic expression is related to the world via the conceptual structure that is mapped to it by the correspondence rules.

However, this language-world relation is not endorsed by Jackendoff. He writes:

> I ... take issue with the naive (and nearly universally accepted) answer that the information language conveys is about the real world.[149]

---

[148] Jackendoff's theory is not quite this simple. He also holds that there are some linguistic - and, therefore, conceptual - structures that do not fall into the function-argument pattern. Ibid., 70-75. However, these are relatively less important in Jackendoff's theory.

[149] Ibid., 24.

Jackendoff's reason for rejecting the "naive" view is based entirely on psychological considerations: viz., the Gestalt psychologists' principle that there is a large gap between sensory input and the structure of experience. The moral of the classic "vase/two faces" figure,[150] according to Jackendoff, is that experience cannot be totally environmental in origin. The nature of experience is largely determined by cognitive processes. Jackendoff expands on this theme in the following passages:

> ... The world as experienced is unavoidably influenced by the nature of the unconscious processes for organizing environmental input. On cannot perceive the 'real world as it is.'...
>
> Such a view, however, seems to compel us to claim that potentially vast areas of our experience are due to the mind's contribution, even though the experience is of things 'out there in the real world.' The only solution to this apparent conflict between theory and common sense is for the theory to include 'out there' as part of the information presented to awareness by the unconscious processes organizing environmental input. That is, 'out-thereness' is as much a mentally supplied attribute as, say, squareness. ...
>
> If indeed the world as experienced owes so much to mental processes of organization, it is crucial for a psychological theory to distinguish carefully between the source of environmental input and the world as experienced. For convenience, I will call the former the real world and the latter the projected world (experienced world or phenomenal world would also be appropriate). ...
>
> I should also make clear that this distinction between the real world and the projected world is not new. Something like it appears at least as early as Kant.[151]

---

[150] For those who are not familiar with this famous pattern, it can be seen either as a vase or two people in silhouette facing each other. It is reproduced ibid., 24.

[151] Ibid., 26, 28, 29.

According to Jackendoff, we do not have direct access to the real world. Rather, the real world impinges on our sense organs to produce environmental information that is processed by our conceptual structures. Such processing gives rise to "projection"; that is, the conceptual structures, given the appropriate environmental information, will be "projected" to give rise to projected entities in the world of experience. Concepts, then, are "about" the projected world, not the real world; and similarly, the expressions of natural language are also "about" the projected world.

Major Ontological Categories

Jackendoff holds that "objects", "events" and other "entities" within our projected worlds are "referents" of expressions in our conceptual structure, and therefore, of expressions in natural language. Individual lexical items in a natural language can "refer" to projected entities, but so can major phrasal categories, including sentences.[152]

Consider, as an example, the sentence:

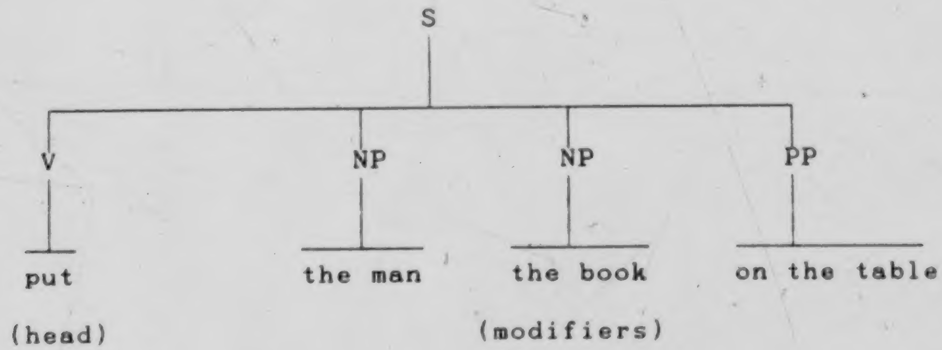
The man put the book on the table[153]

---

[152] This is very different than the traditional semantics of first order logic where only constants, functional expressions, and definite descriptions can play a referential role. In the semantics of first-order logic, sentences are given truth-conditions, and are not considered to be referring expressions. Similarly, many of the sentential constituents that are treated as predicates in the semantics of first order logic are treated as singular referring expressions by Jackendoff. Jackendoff is fully aware of these differences, and considers his analysis superior. See ibid., 57-70 for a comparison of the two approaches, and why Jackendoff rejects the traditional approach.

[153] Ibid., 67-68.

Syntactically, Jackendoff would assign the following phrase structure to this sentence:



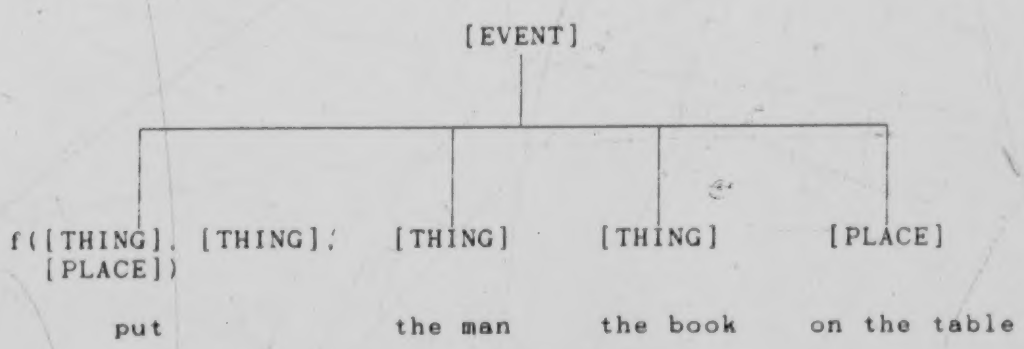
Semantically, the three modifiers are considered zero-place functions that "refer" (I will subsequently drop the scare-quotes) to projected entities that are picked out by means of a set of loosely specified rules defined over semantic markers and which represent the sense, or meaning, of these lexical items.[154] The claim that the two noun phrases have a referential function is not unusual, but the claim that the prepositional phrase 'on the table' is referential will strike many as quite controversial (as will the claim that the sentence is referential). Jackendoff claims that this prepositional phrase refers to a "place". The conceptual constituent associated with the prepositional phrase 'on the table' is characterized, in part, by a primary semantic marker, [PLACE]. Similarly, the conceptual constituents associated with the two noun phrases are characterized, in part, by the primary semantic marker [THING].

---

[154] Jackendoff calls these loosely defined rules "preference rule systems". Ibid., chapter 8.

This sentence has one major phrasal constituent, the sentence itself. The head of this phrasal constituent is the verb 'put', which semantically is treated as a function that takes three arguments: two [THING]s and a [PLACE]. The output of this function will be another conceptual structure whose primary semantic marker is [EVENT].

Thus, the semantic, or conceptual, structure of this sentence mirrors the syntactic structure. A partial representation of the semantic structure is shown below:



This is only a partial representation because the conceptual structures at the nodes actually consist of more than the primary semantic markers shown here. Furthermore, some of the constituents should be broken down further. For example, the phrase 'on the table' has the head 'on', which semantically is considered to be a function that maps [THING]s to [PLACE]s. However, sufficient detail has been presented to convey the general character of Jackendoff's approach.

I have used the phrase 'primary semantic marker' to characterize the main conceptual feature of a conceptual constituent. Jackendoff uses another phrase: 'major ontological category'. There are a fixed, finite number of major ontological categories in our conceptual system, and they play a key role in the semantic analysis of sentences:

> Every major phrasal constituent in the syntax of a sentence corresponds to a conceptual constituent that belongs to one of the major ontological categories.[155]

[THING], [PLACE] and [EVENT] are three of the major ontological categories, and when they are "projected" we experience a (projected) world of things, places and events. There are several other major ontological categories, including [DIRECTION], [PATH], [ACTION], [MANNER], [AMOUNT], and [STATE]. These categories determine the makeup of our projected world; the world that we experience.

> The major ontological categories ... characterize the distinction among the major classes of [projected] entities that we act as thought the [projected] world contains.[156]

An obvious question is: how did Jackendoff come to the conclusion that these concepts, and not some others, are the major ontological categories? The answer is that he determined them by reflecting on certain properties of English.[157] Jackendoff justifies the use of linguistic evidence by invoking the Grammatical Constraint, discussed earlier. I will discuss just one class of evidence that Jackendoff uses, in order to convey the flavor of his approach.

---

[155] *Ibid.*, 67

[156] *Ibid.*, 51.

[157] *Ibid.*, 48-56.

Jackendoff asks us to consider utterances like the following:

(i) I bought that yesterday. (Speaker is pointing)

(ii) You shuffle cards this way. (Demonstrating)

(iii) The fish was this long. (Demonstrating)

Jackendoff argues that (1) the demonstrative terms (underlined) in these utterances can be understood only if they refer to projected entities, and (2) the category of entity referred to can be determined by considering what phrases can be substituted for the demonstrative. The substitution test shows us that the demonstrative in sentence (i) refers to an entity in the ontological category [THING], whereas sentences (ii) and (iii) illustrate the ontological categories [ACTION] and [AMOUNT] respectively.

Jackendoff does not present a satisfactory argument for either point (1) or (2). Since my purpose here is expository, I will pass over this point, however, it should be noted that all the evidence that Jackendoff uses to identify the major ontological categories is linguistic in nature.

The Thematic Relations Hypothesis

According to Jackendoff's theory, the meanings of individual lexical items are represented as n-place functions (possibly zero-place) of conceptual constituents. When a function is provided its arguments it (generally) will refer to some projected entity in the projected world.

Jackendoff devotes one chapter of his book to studying the semantics of spatial terms, including terms of motion and location. Spatial terms are, of course, semantically related. Using a term from Post-Saussurian semantics, we can say that these terms apply to a common "semantic field".[158] In his discussion of spatial terms, Jackendoff tries to demonstrate that the conceptual counterparts of these terms fit together in a formal architecture that has certain properties (which Jackendoff calls "thematic relations"). Another way of putting this is that Jackendoff uses his theory to formalize the semantic field of space and location. (I will not go into the specifics.)

The Thematic Relations Hypothesis is the claim that the conceptual architecture of the spatial semantic field generalizes to other semantic fields. Jackendoff fully endorses this hypothesis:

> The significance of this insight to the present undertaking cannot be overemphasized. It means that in exploring the organization of concepts that, unlike those of [projected] physical space, lack perceptual counterparts, we do not have to start de novo. Rather, we can constrain the possible hypotheses about such concepts by adapting, insofar as possible, the independently motivated algebra of spatial concepts to our new purposes.[159]

I will not go into the details of the structural similarities that Jackendoff sees between the semantics of spatial terms and other semantic fields. It is sufficient to point out that the semantic field of location and motion has a certain conceptual priority, and exerts a significant influence on the structure of many other semantic fields. Jackendoff tries to downplay the idea that there is a pervasive "spatial metaphor" behind all our conceptual structures; preferring to say that "all fields have essentially the same structure"[160] rather than that some fields are structured in terms of others. However, the primacy of the spatial field comes out in the following passage:

---

[158] "Semantic fields" are closely linked to componential analysis. Semantic fields are the "conceptual spaces" that are produced by applying Boolean operations over semantic markers. For details see G. Leech, Semantics, (Penguin, 1974), 96 ff., and especially J. Lyons, Semantics, Volume I, (Cambridge University Press, 1977), 250-261.

[159] R. Jackendoff, Semantics and Cognition, (The M.I.T. Press, 1983), 188.

> If there is any primacy to the spatial field, it si because this field is so strongly supported by nonlinguistic cognition; it is the common ground for the essential faculties of vision, touch and action. From an evolutionary perspective, spatial organization had to exist long before language. One can imagine the development of thematic structure in less concrete fields as a consequence of evolutionary conservativism in cognition - the adaptation of existing structure to new purposes rather than the development of entirely novel mechanisms.[161]

Final Remarks

This completes the overview of Jackendoff's semantic theory. Recall that we began examining this theory because of a complaint that David Lewis raised against Katz and Fodor's theory, viz., that it amounts to nothing more than a translation algorithm. Lewis' complaint is valid, because although Katz and Fodor intended that the semantic markers represent concepts that can be used to pick out things in the world, they did not explain in any detail how this works. Consequently Lewis is correct in pointing out that, so far, conceptual semantics is really just a translation exercise.

---

[160] Ibid., 209.

[161] Ibid., 210.

Jackendoff's theory, on the other hand, is an attempt to follow the program through, and explain how concepts link up to the world. Jackendoff modifies the problem somewhat, claiming that we cannot talk about the real world, only a Kantian "projected" world, but it is clear that his psychological theory is intended to fill the void that Lewis has pointed out.

In the next subsection I will point out the strong similarities between Whorf's Linguistic Relativity Hypothesis and Jackendoff's theory. In the subsection following that I will critically evaluate Jackendoff's theory.

Jackendoff's Conceptual Semantics and Whorf's Thesis

As was pointed out earlier, Whorf wrote during the period in which Bloomfield's influence dominated American linguistics. Semantics, as we have seen, is not something that a clear-thinking Bloomfieldian would engage in, and consequently Whorf broached the subject of semantics (which is essential to his Linguistic Relativity Hypothesis) with a certain hesitancy. Furthermore, a strict Bloomfieldian held that behaviorism is the proper method in psychology, and consequently, Whorf was going against the flow when he proposed that the application of concepts is central to human psychology. Perhaps because of these key differences between his views and those of the dominant linguists of his time, Whorf failed to develop his semantics or his psychology to the point where his Linguistic Relativity Hypothesis is clear and unambiguous.

In any case, it is fairly clear, I think, that Whorf's ideas are much more consonant with the theories of conceptual semantics than with Bloomfieldian linguistics; at least that is what I will argue in this sub-section. In particular, I will show that there are striking similarities between Jackendoff's theory and Whorf's thesis. I will argue that Jackendoff's theory, with a few (significant) modifications, can be regarded as a clarification of Whorf's Linguistic Relativity Hypothesis. Specifically, I will argue that if the elements of Chomskyan nativism are removed from Jackendoff's theory, it is compatible with Whorf's Linguistic Relativity Hypothesis. Diagramatically:

Jackendoff's Theory of Semantics and Cognition - Chomsky's Nativism = Whorf's Linguistic Relativity Hypothesis

In section 1.2 Whorf's argument for the Linguistic Relativity Hypothesis was presented as the following five theses:

1. To think is to employ concepts.
2. Languages have a conceptual structure.
3. Languages differ.
4. Language determines thought.
5. Cross-cultural understanding is difficult/impossible.

These five theses will now be reconsidered in the context of Jackendoff's theory of semantics and cognition.

1. To think is to employ concepts.

Whorf argued that human thought is fundamentally the exercise of concepts, and that without concepts, experience would be amorphous. His view was that the experienced world is, in part, a cognitive construction, and that concepts are the elements from which the mind constructs an experienced world. Finally, Whorf held that there are "cosmic forms": a few key concepts that play a primary role in our mental life and in the construction of an experienced world.

Jackendoff holds similar views. For him as well, cognition is primarily a matter of the exercise of our conceptual system. Like Whorf, he holds that without concepts experience would be amorphous; in fact he states a stronger Kantian doctrine: experience is not possible at all without concepts - experience is "projected" from our conceptual structures when environmental information is provided.

Furthermore, like Whorf, Jackendoff believes that some concepts within our conceptual system are more important than others. Specifically, the "major ontological categories" are Jackendoff's counterpart to Whorf's "cosmic forms". Furthermore, of the major ontological categories, [PLACE]s and [PATH]s (which define the semantic field of space and motion) have a primacy, for their internal logic applies to many other conceptual structures as well (the Thematic Relations Hypothesis).

2. Languages have a conceptual structure.

Whorf held that languages have a conceptual structure, and that this conceptual structure has close ties to the syntactic categories of the language. In particular, Whorf felt that the less overtly marked syntactic categories (which he called "cryptotypes"), are typically correlated with important concepts, i.e., cosmic forms.

As a practitioner of conceptual semantics, Jackendoff also believes that languages have a conceptual structure. However, Jackendoff has gone much farther than Whorf in working out this notion. There are at least three ways in which Jackendoff has gone beyond Whorf in working out this view.

First, Jackendoff has worked out a theory of the internal structure of concepts. Using a decompositional approach that comes from componential analysis, combined with a function-argument framework, Jackendoff has presented a detailed theory of how many of our concepts are made up, and how they can be combined, through conceptual well-formedness rules, to form other concepts.

Secondly, Jackendoff has worked out a theory of how concepts are mapped to language that is both clearer and more detailed than Whorf's. At the lexical level, Jackendoff has provided numerous examples of how to analyze words using his approach. At the level of syntactic categories, he has worked out a very straightforward theory based on the X-Bar theory of categories. Specifically, he holds that all "major phrasal categories" defined in this theory will be associated with a major ontological category, with the head of the phrasal category being related to its modifiers as function to arguments. Whorf held that cosmic forms are most likely to be associated with cryptotypes, i.e., syntactic categories that are not marked in an overt fashion. Jackendoff claims something similar: his theory associates ontological categories not with lexical categories (i.e., parts of speech), which are relatively obvious, but rather to phrasal categories as defined by the X-Bar theory, which are relatively less overt.

Thirdly, Jackendoff has developed his theory of concepts to the point where the ontological consequences are fairly clear. Theorists who treat concepts as real entities fall into two camps: those who view concepts as Platonic entities and those who view concepts as certain types of structures, processes and/or dispositions in the psychological makeup of people (and possibly other animals). Whorf never really makes his position clear, though on balance it is probably fair to say that he endorses the psychological view of concepts. With Jackendoff there is no hesitancy. He clearly views concepts as psychological in nature. Consequently, on his view languages have concepts, and therefore semantic properties, only because languages are internalized by human beings. The internal psychological system of "correspondence rules" (see figure 2.3.4) is what established the connection between syntactic structures and conceptual structures.

3. Languages differ.

Whorf believed that there can be profound differences between languages, both in terms of syntactic structure and in the system of concepts associated with languages. He argued that there are drastic differences between Hopi and Standard Average European (SAE), for example.

Jackendoff, on the other hand, accepts Chomsky's argument that the problem of language acquisition (i.e., how can a child learn something so complex as a language when the information presented is so limited and degenerate?) entails that there must be an innate language acquisition device. This device allows the child to learn languages with a very complex syntax (i.e., unrestricted rewrite systems) within a certain range. No unrestricted-rewrite language outside this range could be learned by a human being, therefore all human languages must share the properties of languages within this range.

Anyone who accepts this Chomskyan argument, as Jackendoff does, tends to concentrate on the commonalties between human languages, rather than the differences. However, two points should be kept in mind. (1) Even if one accepts the Chomskyan argument, there are still significant differences between languages that must be captured in the linguistic descriptions of those languages. Furthermore, the linguistic descriptions should reflect the fact that there is a greater difference between English and Hopi than there is between English and German. (2) It is far from clear that the Chomskyan argument to linguistic nativism should be accepted. Recent theories have suggested that natural languages may only require phrase structure syntaxes after all.[162] If this is the case, then Chomsky's argument - that innate structures must be postulated to explain the gap between the paucity of what is presented and the complexity of what is learned - is weakened accordingly.

---

[162] E.g., G. Gazdar, et. al., Generalized Phrase Structure Grammar, (Harvard University Press, 1985). For an introduction to this theory, see P. Sells, Lectures on Contemporary Syntactic Theories, (Center for the Study of Language and Information, 1985).

In light of these two points, one can imagine modifying Jackendoff's theory such that different syntactic structures are attributed to different languages. That is, the X-Bar theory could be modified so that different syntactic structures are assigned to English and Hopi, for example. Such a modification would not violate the internal logic of the theory; it only requires that we reject Chomsky's argument for nativism, something that a number of writers have advised that we do in any case.[163]

Assuming that two languages, such as English and Hopi, have been assigned significantly different syntaxes, can we then associate them with significantly different conceptual systems? There is no reason why not. As we have seen, Jackendoff used linguistic evidence to determine the "major ontological categories". If we allow that syntactic structure can vary significantly between language, then Jackendoff's own tests for determining conceptual correlates of syntactic structure should generate significantly different results between languages.

In summary, although Jackendoff accepts Chomsky's argument for linguistic nativism, and therefore tends to focus on the commonalties between languages, the Chomskyan premises could be dropped from Jackendoff's theory. This would allow the postulation of significant differences between languages, both in terms of syntactic structure and conceptual content.

---

[163] See, for example, the contributions by H. Putnam and N. Goodman to the "Symposium on Innate Ideas", Boston Studies in the Philosophy of Science, Vol. III, (The Humanities Press, 1968), and B. Derwing, Transformational Grammar as a Theory of Language Acquisiton, (Cambridge University Press, 1973).

4. Language determines thought.

Whorf argued that the language we learn determines the way we think. However, it was noted in Chapter One that Whorf was not particularly clear in making this point. Specifically, it was pointed out that this claim could be interpreted in two ways. Whorf may have intended that the claim be interpreted causally; that is, that it be interpreted as meaning, roughly, that the acquisition of language in one part of the speaker's head will affect the formation of concepts in another part of his head. Alternatively, Whorf may have intended a non-causal interpretation; that is, that one simultaneously learns languages and concepts, and they are just two aspects of one cognitive system in the speaker's head.

Jackendoff's theory (minus the Chomskyan assumptions) is a lot clearer; it unequivocally supports the causal interpretation. Jackendoff holds that conceptual well-formedness rules are innate (recall that his argument for this comes from Fodor, and is independent of the Chomskyan argument for syntactic nativism), but that specific conceptual structures are formed by the rules through a learning process. Jackendoff specifically allows that the learning of syntax and vocabulary can influence the formation of conceptual structures to allow a "certain amount of 'Whorfian' variation in concepts due to linguistic experience."[164] If we drop the Chomskyan assumptions from Jackendoff's theory, then the theory will recognize a greater amount of variation between languages, and consequently, the amount of "Whorfian" variation in concepts will be greater.

---

[164] R. Jackendoff, Semantics and Cognition, (The M.I.T. Press, 1983), 242.

Jackendoff's theory definitely supports the causal interpretation of the claim that language affects thought, as can be seen by examining figure 2.3.4. This diagram shows that in Jackendoff's view, the linguistic system and the conceptual system ares considered separate sets of mechanisms in the speaker's head. The learning of language, for example, the use of demonstratives, is stored in the linguistic system. This learning may then cause changes to the conceptual system - or to put it more acceptably, the conceptual system may be pressured into changing in order to accommodate the learning. The learning of demonstratives is especially relevant, because it will be recalled that Jackendoff stated that demonstratives are a clue to the major ontological categories. It is not too much of a stretch to argue that if Chomskyan nativism is subtracted from Jackendoff's theory, then it will imply that the learning of demonstratives in a natural language causes the child to develop the major ontological categories associated with that language.

To put the point more generally, Jackendoff himself used only linguistic evidence to determine the major ontological categories, so it could also be argued that the child has only linguistic evidence on which to formulate his ontological categories. If the linguistic evidence available in different speech communities varies widely, then so will the ontological categories - or so it could be hypothesized.

5. **Cross-cultural understanding is difficult/impossible.**

Whorf argued that cross-cultural communication is very difficult if the languages of the two cultures are significantly different.

Jackendoff's theory (minus Chomskyan nativism) also implies this. Jackendoff holds that all people have an innate system of conceptual well-formedness rules. The actual concepts formed will be a function of the environmental information to which the person is exposed, including information about the natural language learned by that person. If, as suggested, we drop - or at least soften - Chomsky's views on innate syntax from Jackendoff's theory, then the effect of language on the formation of conceptual structures will be increased. It follows, then, that in order to "calibrate" the conceptual systems of speakers of two radically different languages, it will be necessary to have a sophisticated understanding of the syntactic and lexical structures of each language, together with an understanding of how those linguistic structures interact with the conceptual well-formedness rules to produce conceptual structures.

Whorf's Linguistic Relativity Hypothesis did not conform well to the Bloomfieldian approach to the study of language. However, as has been demonstrated, it fits in quite well with the tradition of conceptual semantics; a tradition that began with Saussure and was further developed by the attempt to incorporate Saussurian semantics into generative grammar. The key modification that must be made to such theories is to deny the strong syntactic nativism that such theories usually assume. Unfortunately for Whorf's intellectual heritage, however, there are serious problems with the program of conceptual semantics.

Problems with Conceptual Semantics

Whorf's Linguistic Relativity Hypothesis fits well with the conceptual semantics approach. In fact, the hypothesis is improved in clarity within the context of a theory like Jackendoff's, for Jackendoff provides a much more detailed account of many of the issues that are critical to the Linguistic Relativity Hypothesis. However, we will not be able to retain this reading of the Linguistic Relativity Hypothesis, for there are serious problems with conceptual semantics. There are two problems I want to raise:

Problem 1 - Kantianism

Problem 2 - "Meaning ain't in the head"

These two problems will now be discussed in turn.

Problem 1 - Kantianism

David Lewis, it will be recalled, complained that Katz and Fodor's theory is just a system of translation, not a semantic theory. Lewis insists, and rightly so, that a semantic theory must describe how a language is related to the world; it must describe the sense in which the language is "about" the constituents of the world. Katz and Fodor's theory fails to do this; it merely provides a translation algorithm between the natural language being analyzed and another language that we can call "Markerese".

It was pointed out that conceptual semantics could be saved from Lewis' complaint if it were supplemented with a theory that takes the Markerese symbols to stand for internalized concepts, where these internalized concepts are "about" the constituents of the world. In other words, to avoid Lewis' complaint it is necessary to develop a psychological theory where Markerese is taken to be an internal system of conceptual structures that have true semantic properties in Lewis' sense.

Jackendoff's theory was examined as an example of such an approach. However, there is a significant difference between what we were after and what Jackendoff provided. What we were after was an account of how the internalized conceptual system has semantic relations to the world. What Jackendoff gave us was a theory of how the internalized conceptual system has semantic relations to a "projected world". Jackendoff, it will be recalled, denies that our concepts can directly apply to the real world. I called this aspect of Jackendoff's work his "Kantianism".

Given that our pre-theoretical intuitions are that our language and our thoughts are about the real world, not a projected world, Jackendoff's Kantianism deserves critical examination. I will argue that Jackendoff's Kantian position should be rejected. This interlude will pave the way for the discussion of the second problem with conceptual semantics.

Jackendoff's argument for the claim that we only have access to a projected world is based entirely on observations about the nature of perception. Specifically, he rehearses points made by the Gestalt psychologists to the effect that the sensory stimulus giving rise to perception contains insufficient information to account for the characteristics of what is perceived. To take the classic vase/two faces example, clearly the sensory stimulus is unchanged in the two interpretations. Consequently, something must be added to the sensory stimulus to produce the vase interpretation, and something different must be added to produce the two faces interpretation. According to Jackendoff, such examples show that:

> ... what one sees cannot be solely environmental in origin, since the figures are imbued with organization that is not there in any physical sense... The organization, which involves both segmentation of the environmental input and individuation of disparate parts, must be part of the mind's own encoding of the environmental input...
>
> Thus, the world as experience is unavoidably influenced by the nature of the unconscious processes for organizing environmental input. One cannot perceive the 'real world as it is'.[165]

Jackendoff's argument is an epistemological argument, and it will help to place it in the larger geography of the epistemology of perception. It is common to categorize philosophical theories of perception into three main categories: First, "direct realism" holds, at a minimum, that we are directly aware of physical objects and that at least some of the properties of physical objects that we perceive are retained by those objects when we are not perceiving them. Secondly, "indirect realism" holds that there are objects that exist independently of our perception of them, but that we cannot perceive them directly. Rather, we are directly aware only of internal non-physical objects. Thirdly, "phenomenalism" agrees with indirect realism that we are directly aware only of internal non-physical objects, but disagrees with both direct and indirect realism that there are objects that exist independently of the objects of direct awareness. Phenomenalism holds that there is only a "phenomenal world", and that the objects that we take to be real (independently existing) objects are actually constructions from the elements of the phenomenal world.[166]

---

[165] R. Jackendoff, Semantics and Cognition, (The M.I.T. Press, 1983), 25, 26.

In terms of this typology, Jackendoff is an indirect realist, for he holds that we can only be directly aware of a "projected world" (he says that the term "phenomenal world" can be substituted) and never the real world as it is. Jackendoff's argument in favor of indirect realism is a form of the "argument from illusion", which is just one of the arguments that has been historically presented in favor of indirect realism.[167] An "argument from illusion" has the following form: my perception of object O has characteristic C; however, the sensory stimulation that O produces in me could not possibly have characteristic C; therefore I must not be directly aware of O, but rather, some intermediary that has the characteristic C.

---

[166] This typology of theories is discussed in greater detail in J. Dancy, Introduction to Contemporary Epistemology, (Basil Blackwell, 1985), chapter 10.

[167] Ibid., 152-154.

I reject indirect realism in favor of direct realism. A defense of this claim would be a major undertaking, which will not be attempted here. Instead, I offer a brief sketch of why I endorse direct realism.

First, there are serious problems with the notion that we are directly aware, not of the world, but of an intermediary. What is the ontological status of the intermediary objects of which we are directly aware? Are they made up of some sort of Cartesian "mental substance"? The problems with dualism are well known and formidable.[168] Another problem that has been frequently pointed out is that it is very difficult for the indirect realist to distinguish his position from phenomenalism. If it is held that we can never be directly aware of any of the properties of the real world other than its mere existence, what good reason do we have to hold that the real world even exists?[169] This is the charge that the phenomenalists Berkeley and Hegel made against the indirect realists Locke and Kant, respectively. Historically there has always been a great tendency for indirect realism to collapse into phenomenalism. However, phenomenalism faces grave difficulties. The greatest efforts in developing a phenomenalist philosophy were in the early decades of the twentieth century, through the work of theorists like Russell and Carnap. However, these attempts are generally regarded as failures, and it is generally felt that any form of phenomenalism, and the reductive analyses that it requires, is faced with insurmountable objections.[170]

---

[168] See, for example, P. M. Churchland, Matter and Consciousness, (The M.I.T. Press, 1984), 18-21.

[169] J. Dancy, Introduction to Contemporary Epistemology, (Basil Blackwell, 1985), 164-165.

Secondly, the issue that prompted Jackendoff to endorse indirect realism can be accommodated by the direct realist. The problem that Jackendoff pointed out is that there can be a gap between the environmental information that is available to the perceiver and the information that is available in the perception. However, if we take the having of perceptions to be a variety of belief formation, then the gap can be understood as a special case of the underdetermination of theory (sets of beliefs) by evidence.[171] Underdetermination of beliefs by evidence does not entail that we are directly aware only of "intermediaries", it merely entails that there is room for empirically unconditioned embellishments in our system of beliefs.[172]

---

[170] R. J. Hirst, "Phenomenalism", in The Encyclopedia of Philosophy, Volume VI, ed. by P. Edwards, (Macmillan Publishing, 1967), 130-135.

[171] Whether perception can be totally understood as a type of belief is controversial. However, Dancy argues that even if perception is considered to be a "mixture" of sensation and belief, the belief component is enough to refute the indirect realists "argument from illusion". See J. Dancy, Introduction to Contemporary Epistemology, (Basil Blackwell, 1985), 169-173.

[172] For a development of this idea that the gap between environmental information and what we believe can be accounted for in a realistic framework by invoking underdetermination, see W. V. O. Quine, Word and Object, (The M.I.T. Press, 1960), chapter 1. Quine specifically addresses the findings of the Gestalt psychologists, which Jackendoff used to argue for indirect realism, in "Epistemology Naturalized", in Ontological Relativity and Other Essays, (Columbia University Press, 1969), 84-85, and argues that they can easily be handled within his realist framework.

Given that there are serious problems with Jackendoff's Kantianism, I suggest that we excise that aspect from his theory. What we are left with is a realistic theory in which people are said to have conceptual systems in their heads, and these conceptual systems refer to objects, events, etc., in the real world.[173] Furthermore, according to this realistic version of Jackendoff's theory, the expressions of natural language refer to the constituents of the real world in virtue of their correlation with the conceptual structures in people's heads. Such a theory will allow the advocate of conceptual semantics to avoid David Lewis' objections.

However, it leads directly to the second objection to conceptual semantics.

Problem 2 - "Meaning ain't in the head."

We have argued that once Jackendoff's theory is cleansed of its objectionable Kantianism it can account for the referential properties of linguistic expressions by the hypothesis that linguistic expressions are correlated with conceptual structures in people's heads, and that these conceptual structures have referential properties. However, Hilary Putnam has presented a powerful and influential argument that nothing in the heads of people could possibly have the reference determining properties that we are looking for.[174]

---

[173] Jackendoff would object to this move, for he holds that certain aspects of his theory, for example, the Thematic Relations Hypothesis, are "profoundly incompatible" with a realistic framework. See Semantics and Cognition, (The M.I.T. Press, 1983), 208-209. However, in light of the other objections I will be presenting against conceptual semantics, I choose to overlook this possible conflict within Jackendoff's theory. It should perhaps be mentioned that most conceptual semantic theories do not take the Kantian turn that Jackendoff's does, and are therefore compatible with a realistic framework. See, for example, J. A. Fodor, The Language of Thought, (Thomas Y. Crowell, 1975).

Putnam argues that since the Middle Ages, writers concerned with the issue of "linguistic meaning" have made a number of assumptions that, when taken together, Putnam calls the "traditional theory". For the purposes of his argument, Putnam focuses on the semantics of natural kind terms, such as 'water', 'tiger' and 'gold'.

The traditional theory holds, first of all, that terms have both an extension and an intension. The extension is, roughly, the set of objects that the term is true of. (Roughly, because natural kind terms divide into two types; mass terms, like 'water' and 'gold', and sortal terms, like 'tiger'. While the extension of sortal terms can be straightforwardly viewed as sets of objects, the extension of mass terms cannot. However, this complication will be ignored here, as it is in Putnam's paper.) The intension is the "meaning" of the term, or the "concept" that is associated with it. The distinction between extension and intension comes from the recognition that two terms, say 'chordate' (creature with a heart) and 'renate' (creature with kidneys), may have the same extension, but intuitively they differ in meaning. Traditionally, intensions are thought to "determine" extensions in the sense that the intension is a set of necessary and sufficient conditions for inclusion in the extension.

---

[174] H. Putnam, "The Meaning of 'Meaning'", Language, Mind and Knowledge: Minnesota Studies in the Philosophy of Science, Volume VII, ed. by K. Gunderson, (University of Minneapolis Press, 1975).

The traditional theory holds that people understand terms by knowing their intensions. Many theorists have held that intensions are inherently mental. Others, such as Frege and Carnap, have taken a Platonic view of intensions, arguing against psychologistic reductions. However, even these theorists have said that an individual comes to understand a term by "grasping" its intension, so they are, in effect, arguing that when a person understands the natural kind terms of a language, he has psychological states that are in a one-to-one correspondence with the intensions of the natural kind terms. (For this reason, Putnam states that "the whole psychologism/Platonism issue appears somewhat a tempest in a teapot, as far as meaning-theory is concerned."[175]) Furthermore,

> ... when traditional philosophers talked about psychological states (or 'mental' states), they made an assumption which we may call the assumption of methodological solipsism. This assumption is the assumption that no psychological state, properly so called, presupposes the existence of any individual other than the subject to whom that state is ascribed.[176]

According to the assumption of methodological solipsism, knowledge, for example, cannot be (solely) a psychological state, for a condition of knowing p is that p must be true, and that will, in general, depend on circumstances external to the individual who is doing the knowing. The purpose of the assumption of methodological solipsism, then, is to ensure that matters psychological are not confused with matters that are external to the psychological subject.[177]

---

[175] Ibid., 138.

[176] Ibid., 136.

The assumption of methodological solipsism means that psychological states will have to be carefully described in order to retain the assumption of methodological solipsism. For example, it is natural to assume that one's beliefs have truth-values; that is, one can be right or wrong about things. Now suppose that you have prepared a molecule for molecule replica of me, and then kidnaped me and put the replica in my place. According to the assumption of methodological solipsism, the replica would be psychologically identical to me (assuming materialism to be true). But notice my replica "believes" that he once slept in Tuktoyaktuk. This "belief" is false for him, but true for me. Consequently, in order to maintain the assumption of methodological solipsism we will have to revise our concept of belief so that it does not have the semantic property of carrying a truth-value. We will have to distinguish what Putnam calls "narrow" psychological states from "wide" psychological states. An example of a wide psychological state is our ordinary concept of belief, which as the replica example shows, involves relations between the believer and external objects or events. A example of a narrow psychological state would be whatever the replica and I have in common, exclusive of our relations with other objects or events, when we are both "believing" that we once slept in Tuktoyaktuk.

---

[177] Stephen Stitch uses the term "the principle of psychological autonomy" to describe what Putnam calls "methodological solipsism". Stitch stresses what Putnam does not, namely, that this principle "serves as a fundamental regulative principle for much of contemporary psychological theorizing". See S. Stitch, "Autonomous Psychology and the Belief-Desire Hypothesis", The Monist, 1978, 61, 573-591.

To summarize, Putnam argues that the traditional theory of meaning involves the following three assumptions:

> (M) Methodological Solipsism - No psychological state, properly so called, presupposes the existence of any individual other than the subject to whom that state is ascribed. In other words, every psychological state, properly so called, is a "narrow" psychological state.
>
> (I) Knowing the meaning of a term is just a matter of being in a certain (narrow) psychological state.
>
> (II) The meaning of a term (in the sense of "intension") determines its extension (in the sense that sameness of intension entails sameness of extension).[178]

Together, these three principles imply the following:

> (P) A competent speaker's knowledge of the meaning of a term is a (narrow) psychological state that determines the extension of the term.

But Putnam has developed some counter-examples to show that (P) is not true, and therefore at least one of the three premises will have to be given up.[179] Let us consider two of Putnam's counter-examples.

---

[178] H. Putnam, "The Meaning of 'Meaning'", Language, Mind and Knowledge: Minnesota Studies in the Philosophy of Science, Volume VII, ed. by K. Gunderson, (University of Minneapolis Press, 1975), 135-136.

[179] Putnam only considers the possibility of giving up either (I) or (II). However a third possibility is to give up (M), and engage in what Fodor has called "naturalistic psychology", i.e., a psychology that includes in its domain many of the organism's relations to its environment. This possibility is discussed, and rejected, in J. A. Fodor, "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology", The Behavioral and Brain Sciences, 1980, 3, reprinted in J. A. Fodor, Representations, (The M.I.T. Press, 1981), 228. However, I will be defending this third option in the next chapter.

(1) 'Beech' and 'Elm'

Putnam says that he is not able to distinguish beeches from elms, and therefore his internal representation of the meanings of 'beech' and 'elm' are indistinguishable. Therefore, if (P) were true, then when Putnam uses either 'beech' or 'elm' he would not be capable of successfully referring to either species of tree, but only a roughly defined class of deciduous trees (minus the species of deciduous trees that he can recognize) in general. However, this is contradicted by the facts of how language is actually used. If Putnam were, for example, to ask for an elm sapling from a tree nursery, and was given a beech instead, he would be within his rights to ask for an exchange if the mistake were later pointed out to him. The proprietor of the tree nursery could not successfully argue in small claims court that he gave Putnam what Putnam asked for on the grounds that in Putnam's idiolect a beech falls within the extension of 'elm'.

What this example shows is that the extension of terms is socially determined, not individually determined. In general there is what Putnam calls a sociolinguistic division of labor.

> HYPOTHESIS OF THE UNIVERSALITY OF THE DIVISION OF LINGUISTIC LABOR:
>
> Every linguistic community exemplifies the sort of division of linguistic labor just described, that is, possesses at least some terms whose associated "criteria" are known only to a subset of the speakers who acquire the terms, and whose use by the other speakers depends on a structured cooperation between them and the speakers and the relevant subsets.[180]

---

[180] H. Putnam, "The Meaning of 'Meaning'", Language, Mind and Knowledge: Minnesota Studies in the Philosophy of Science, Volume VII, ed. by K. Gunderson, (University of Minneapolis Press, 1975), 145-146.

In other words, the court would rule in favor of Putnam, because the extension of the term 'elm' is held to be determined by a subset of "experts" who can distinguish elms from beeches. That is, there is a general social expectation that when either Putnam or the proprietor or anyone else use the term 'elm', its extension is determined by what the experts think, not what a particular layman thinks.

Putnam's point is that our actual use of terms (e.g., the outcome of the hypothetical court case) conforms with the hypothesis of the division of linguistic labor and contradicts (P). However, when we theorize about language, there has been a strong tendency to (incorrectly) think about it as entirely a matter of individual psychology.

(2) 'Water' on Earth and Twin Earth

Another counter-example to (P) is Putnam's famous Twin Earth example, which has received a huge amount of discussion since it was published. Putnam asks us to imagine another planet, Twin Earth, which is exactly like earth except that what they call 'water' (in the Twin English speaking community) is not H_2O, but XYZ. XYZ is exactly like H_2O in its phenomenal properties, i.e., it is colorless, quenches thirst, etc., but differs in its chemical makeup.

Putnam asks us to think back to 1750, before chemical theory was developed on either planet. Then every English speaking person on Earth will have a double on Twin Earth, and these pairs of people will have identical psychological states when they use the term 'water'. But, Putnam claims, the extensions of their terms differ. The Earthlings are referring to H_2O, whereas the Twin Earthlings are referring to XYZ. Therefore, what is in the head does not determine extension.

The traditional theory assumes that the extension of these terms is a matter of individual psychology. Putnam argues that the way that extension is really determined is more complex. Natural kind terms have to be understood as theoretical terms. These terms are introduced through ostensive acts - at some point in the history of language the term 'water' was applied to some sample body of water. The intention behind the ostension is that the term should be used to apply to all other bodies of matter that are the same as the sample body used in the original act of ostension. What counts as the "same" is a theoretical issue, and criteria of "sameness" may not be available at the time of the ostensive act. Thus 'water' refers to H_2O on Earth and to XYZ on Twin Earth because the sample bodies that were used in the initial ostensive acts are different kinds of matter.

What the Twin Earth story shows, according to Putnam, is that the extensions of many terms are influenced by what he calls "indexicality", that is, the brute facts about what particular sorts of things people happened to be pointing at during their ostensive acts.

Putnam sums up the relevance of his two counter-examples to the traditional theory in the following passage:

> If there is a reason for both learned and lay opinion having gone so far astray with respect to a topic which deals, after all, with matters which are in everyone's experience, matters concerning which we have, if we shed preconceptions, pretty clear intuitions, it must be connected to the fact that the

> grotesquely mistaken views of language which are and always have been current reflect two specific and very central philosophical tendencies: the tendency to treat cognition as a purely individual matter and the tendency to ignore the world, insofar as it consists of more than the individual's "observations". Ignoring the division of linguistic labor is ignoring the social dimension of cognition; ignoring what we have been calling indexicality of most words is ignoring the contribution of the environment. Traditional philosophy of language, like much traditional philosophy, leaves out other people and the world; a better philosophy and a better science of language must encompass both.[181]

The counter-examples that Putnam has developed show that (P) cannot be maintained (and, more generally, constitute a devastating critique of the assumptions of Saussurian structuralism). If meaning determines extension, then "meaning just ain't in the head". What are the implications?

It means that we will have to give up at least one of the premises behind (P). Which one we give up is perhaps a matter of taste, for giving up any one will violate some of our intuitions. (Our intuitions about meaning are in bad shape, according to Putnam.) Putnam chooses to give up (I). He continues to hold that meaning determines extension, but meaning is only partly psychological; it is also sociological, and also has a lot to do with ostension.

So if meaning ain't in the head, what is in the head that determines how individuals use terms? It is possible to take a soft line or a hard line. The soft line is that we when we know a term we have at least part of a meaning in our head. The hard line is that we have all sorts of beliefs in our head, but no

[181] Ibid., 193.

part of this belief system can be sensibly identified as representing a partial word-meaning. Putnam argues for the soft line. I disagree, and will argue for the hard line.

As stated above, Putnam argues that we should retain the idea that meaning determines extension, but give up on the idea that meaning is in the head. Putnam proposes a "normal form" for the description of meanings, with the following example being the meaning of 'water':[182]

Syntactic Markers	Semantic Markers	Stereotype	Extension
mass noun, concrete	natural kind, liquid	colorless, transparent, tasteless, thirst-quenching, etc.	H_2O (give or take impurities)

The syntactic and semantic markers are similar to what we find in Katz and Fodor's theory, which the exception that Putnam typically assigns fewer semantic markers to a term than Katz and Fodor would. The "stereotype" consists a set of conditions associated with a term, except that these conditions are generally insufficient to determine extensions uniquely, and moreover, they are "defeasible". That is, if it were discovered, for example, that up until now we have only been exposed to slightly abnormal water, and that normal water has a distinctive taste, then we would drop the "tasteless" feature from our stereotype.[183] Finally, the extension is a description of what the term actually refers to.[184]

---

[182] Ibid., 191.

Putnam holds that English speakers, at least the ones who have learned the term 'water', have internalized everything about the meaning except the extension. Thus the 1750 inhabitants of Earth and Twin Earth have identical psychological states (consisting of an amalgam of syntactic markers, semantic markers, and a stereotype). However, to determine the extension of the term on the two planets we will also have to study the world (specifically, the makeup of what is called 'water' on each planet, as well as the division of linguistic labor in the English and Twin English speech-communities).

The reader may now be thinking that Putnam's critique of the "traditional" theory of meaning, although very important in its own right, is not relevant to our topic, which is the question of the validity of conceptual semantics as a theoretical foundation for the Linguistic Relativity Hypothesis. For Putnam allows, after all, that terms have meanings associated with them, and the intensional (conceptual) component of those meanings are internalized by language users. This is all we need to construe

[183] H. Putnam, "Is Semantics Possible?", in H. Kiefer and M. Munitz, Languages, Belief and Metaphysics, (The State University of New York Press, 1970).

[184] Note that Putnam's view of semantics is similar to Bloomfield's insofar as they both believe that the extensions of referring terms are determined by the advances of science, and thus our ability to specify meanings may be hampered by the current state of scientific knowledge. For example, Putnam points out that we are currently not in a position to give a precise description of the extension of 'dog' or 'cow'.

Jackendoff's theory (cleansed of its Chomskyan and Kantian elements) as a basis for the Linguistic Relativity Hypothesis.

However, this conclusion follows only if we agree with Putnam's "soft line" that language users internalize everything in his normal form "meanings", with the exception of the extension. I disagree, and will argue for the "hard line" that there is nothing in a speaker's head that can sensibly be identified as his knowledge of the meaning of a term as opposed to his general knowledge of the world.

Putnam's proposal suggests that the beliefs of a normal speaker of English can be broken into two fundamentally different categories. Consider the following two beliefs:

(1) The belief that 'water' refers to a kind of liquid that is colorless, etc.

(2) The belief that there is water in the Pacific Ocean.

Belief (1) is, according to Putnam's proposal, a meaning-characterizing belief, that is, one knows the meaning of 'water' in virtue of holding this belief. Belief (2), on the other hand, is a belief about the world, not a belief about language. If one failed to hold belief (2) it would not imply a defective grasp of the meaning of 'water', whereas failure to hold belief (1) would constitute a failure of semantic competence. Putnam's proposal, then, suggests that each speaker of English has an internalized "dictionary", consisting of beliefs like (1), and an internalized "encyclopedia", consisting of beliefs like (2).

The problem with Putnam's proposal is that it is impossible to maintain a sharp distinction between dictionary knowledge and encyclopedia knowledge. Note that the presence of a mentioned word (e.g., 'water') does not serve as a criterion of dictionary knowledge. Consider the following:

(3) The belief that water is colorless.

(4) The belief that the liquid in the Pacific Ocean is called 'water'.

Although (3) does not contain a mentioned word, it is, on Putnam's view, true "in virtue of meaning" (although it should be pointed out that Putnam argues, against the traditional doctrine of analyticity, that beliefs like (3) are always defeasible). In other words, Putnam would hold that (3) is true in virtue of dictionary knowledge, not encyclopedia knowledge. (4), on the other hand, mentions the word 'water', but is not true in virtue of meaning. It belongs to the category of encyclopedia knowledge. Consequently, the presence of a mentioned word cannot serve as a criterion to distinguish dictionary knowledge from encyclopedia knowledge.

What does distinguish dictionary knowledge from encyclopedia knowledge? One approach might be to argue, along with Saussurians like Leech, that those beliefs characterized by analyticity are dictionary entries, whereas all non-analytic beliefs are encyclopedia entries. The problem with this approach, as Quine has argued,[185] is that it is not at all clear how to give a non-circular account of analyticity. Putnam agrees with Quine's pessimism about analyticity,[186] and therefore does not feel that it can be used as a basis to distinguish dictionary knowledge from encyclopedia knowledge. Instead, the criterion that Putnam uses to identify dictionary knowledge is that it is the minimal knowledge required to impute communicative competence to an individual. For example, the dictionary knowledge associated with the term 'water' is that minimal knowledge that a person must possess to avoid reactions like "he has no idea what the term 'water' means" from other members of his speech community. A person who points to a forest and asks "Is that water?" lacks that minimal knowledge and would be considered by his peers to not know the meaning of 'water'.

---

[185] W. V. O. Quine, "Two Dogmas of Empiricism", in From A Logical Point of View, (Harper and Row, 1961).

[186] Rather, Putnam agrees for the most part. Unlike Quine, Putnam feels that there is a limited class of statements that are clearly analytic. See H. Putnam, "The Analytic and the

The problem with this criterion is that it is vague to the point of uselessness. On the one hand, a person may fail to know the Putnamian stereotype associated with a word, and still be judged to have semantic competence. I believe that Putnam's own 'beech'-'elm' example shows exactly that. Putnam claims that he has a defective understanding of the stereotypes of these two terms - he only knows what is common to both of them - but I for one would not judge him as thereby suffering from a serious communicative deficiency.

On the other hand, someone might have a perfect grasp of the Putnamian stereotype of a term but his collateral knowledge may be so defective that he also elicits judgments of incompetence

Synthetic", in H. Fiegl and G. Maxwell, eds., Minnesota Studies in the Philosophy of Science, III, (University of Minnesota Press, 1962). However, this limited endorsement of analyticity is not capable of marking the distinction between dictionary knowledge and encyclopedia knowledge in Putnam's view, simply because only a limited set of terms, e.g., 'bachelor', generate analytic sentences.

from his peers. Suppose, for example, that someone holds all the beliefs associated with the Putnamian stereotype of 'water', but in addition, he believes that water is instantly lethal when consumed or put in contact with the body. Such a person, I submit, would be judged by his peers to have a "problem" with the term 'water'.

Putnam's criterion fails to pick out a class of distinctively "semantic" information (the internal dictionary) that can be clearly distinguished from information about the world (the internal encyclopedia). Nor would any other criterion fare any better, for the following reason: Each of us has a number of beliefs about water, from the relatively "central" belief that it is a kind of colorless liquid, to somewhat less central belief that water is necessary for human life, to the highly contingent belief that one has drunk a glass of water in the last hour. We also believe, if we speak English, that 'water' refers to water. From this it follows (by simple deduction) that one also believes that 'water' refers to a kind of colorless liquid, that 'water' refers to something that is necessary for human life, and that 'water' refers to something that one has drunk a glassful of in the last hour. Because of the deductive relations that hold between our beliefs, we acquire beliefs about the content of terms at exactly the same rate that we acquire beliefs about the world. And just as we cannot sharply distinguish "central" facts about the world from those that are relatively "contingent", we also cannot sharply distinguish "central" facts about the meaning of terms from those that are relatively "contingent". Consequently it is entirely artificial to assume, as Putnam does, that there is some basic set of beliefs associated with terms that are necessary and sufficient for semantic competence over those terms.

Summary

Conceptual semantics is an approach to language that originated with Saussure and is currently being revived in the work of people like Ray Jackendoff.[187] This work has strong affinities with Whorf's work, and seems to express the theoretical framework that Whorf was edging toward, but perhaps was reluctant to express because of the dominance of Bloomfield's anti-mentalism at the time. However, we have seen that this approach suffers from a fundamental weakness, namely, it is not capable of accounting for the basic semantic relation, viz., the relation between language and the world. Even granting this weakness one might still hope that conceptual semantics at least provides a model of semantic competence at the level of the individual speaker. Putnam, for example, argues that conceptual semantics will give us at least this much. However, serious doubts were expressed regarding even this possibility. It does not appear to be possible to clearly distinguish an individual speaker's knowledge of language from his knowledge of the world, and this distinction is essential to Putnam's program.

---

[187] For other examples of writers who are currently advocating a "conceptual semantics" approach, see the papers collected in U. Eco, M. Santambrogio, and P. Violi, eds. Meaning and Mental Representation, (Indiana University Press, 1988), especially the papers by Jackendoff, Johnson-Laird, and Lakoff.

Conceptual semantics is beset with difficulties. It is time for a fresh start.

2.4 TRUTH-CONDITIONAL SEMANTICS

In the 1930s, the Polish logician Alfred Tarski developed a method of formally defining the predicate 'is true in L', where L is a certain type of artificial language.[188] Tarski's work is extremely important in the philosophy of mathematics. Prior to Tarski's work one of the popular philosophies of mathematics, advocated by David Hilbert, was formalism, which held that mathematical truth should be identified with provability in a system. Provability is the result of the mechanical movement of uninterpreted symbols; it is, in effect, a "syntactic" notion. But what Tarski did was to provide an independent notion of mathematical truth; a "semantic" notion, that defined the truth of each sentence in terms of a domain of objects that the language's referring terms and predicates are "about". Later Kurt Gödel proved his famous incompleteness theorems, which in effect showed that it is impossible to construct a formal language and a set of deductive rules that is both consistent and capable of proving the true sentences of arithmetic, where the true sentences are independently characterized by means of a Tarskian truth definition. In other words, Tarksi showed that truth is conceptually distinct from provability, and Gödel showed that in at least some mathematical languages, truth and provability cannot even be coextensive.[189] The formalist philosophy of mathematics was rejected as a result of these developments.

---

[188] A. Tarski, "The Concept of Truth in Formalized Languages", in A. Tarski, Logic, Semantics, Metamathematics, (Oxford University Press, 1956). For a more informal account see A. Tarski, "The Semantic Conception of Truth", Philosophy and Phenomenological Research, 1944, 4, 341-376.

Tarski intended his theory of truth to be applied in the field of metamathematics, where it contributed to the downfall of formalism. Tarksi did not propose that his methods be applied to the study of natural languages. However, since his work was introduced, a number of other writers have sought to extend the scope of Tarksi's method of constructing truth definitions to ever more general problems of linguistic description. Carnap made some preliminary steps towards applying Tarskian methods on languages that, although still artificial, talk about more than mathematical objects. Others who contributed toward a generalization of Tarski's methods to non-mathematical languages included Richard Martin, John Kemeny, and Alonzo Church.[190] It was not until the 1960s, however, that the idea of using Tarskian

[189] This is oversimplified. Gentzen later developed a non-finitistic method of proving the consistency of arithmetic, but this method cannot be represented within arithmetic, therefore Gödel's results still stand. Also, because the formalist philosophy of mathematics rejected non-finitistic methods, Gentzen's work cannot save that approach to metamathematics. See E. Nagel and J. Newman, Gödel's Proof, (New York University Press, 1958), 96-97.

[190] For a review of developments up to about 1960, see R. Rogers, "A Survey of Formal Semantics", Synthése, 1963, 15, 17-56.

truth-definitions as a theoretical framework for the study of natural languages was seriously proposed. Donald Davidson's classic paper "Truth and Meaning" is perhaps the most well-known and influential article recommending this approach, but other writers, notably Richard Montague, have advanced similar claims.[191]

In this section I will argue that truth-conditional semantics is a promising approach for the study of natural language. It is not entirely free from objections, and significant areas of natural language present formidable problems for the approach, but I will argue that it is our best bet for the study of natural language. Accordingly, the Linguistic Relativity Hypothesis must be evaluated within the context of this approach.

Two mature statements of the truth-conditional program for the study of natural langauges are Donald Davidson's "Radical Interpretation" and David Lewis' "Languages and Language".[192]

---

[191] D. Davidson., "Truth and Meaning", Synthése, 1967, 17, 304-323. This paper, along with several other important papers relating to this proposal are collected in D. Davidson, Inquiries into Truth and Interpretation, (Oxford University Press, 1984). Montague's key papers are in R. Montague, Formal Philosophy, (Yale University Press, 1974). Two important collections of papers on the application of Tarskian methods to natural languages are D. Davidson and G. Harman, eds., The Logic of Grammar, (Dickenson, 1975), and G. Evans and J. McDowell, Truth and Meaning, (Oxford University Press, 1976). Two book length studies of the program are M. Platts, Ways of Meaning, (Routledge and Kegan Paul, 1979) and W. Lycan, Logical Form in Natural Language, (The M.I.T. Press, 1984).

[192] D. Davidson, "Radical Interpretation", Dialectica, 1973, 27, 313-328. D. Lewis, "Languages and Language, in K. Gunderson, ed., Language, Mind, and Knowledge: Volume VII, Minnesota Studies in the Philosophy of Science, (University of Minnesota Press, 1975).

While Davidson and Lewis are not in agreement on all issues, they share the following four assumptions - assumptions that are definitive of the truth-conditional program.

1. Naturalism about people and their environments
2. Linguistic behavior as a species of action
3. Linguistic Platonism
4. Tarskian truth definitions as a framework for semantics

These four assumptions will now be discussed individually.

Naturalism about people and their environments

The first assumption is that the starting point for the study of natural language is the recognition of a speech-community, that is, a group of people who communicate by means of a common language. Moreover, this community lives in an environment of physical objects. The student of language must assume the existence of people and their environments (although there is room for debate on how "people" and "environments" are to be described).

This assumption was shared by Bloomfield, as we saw previously. However, this assumption was not shared by the program of "conceptual semantics". Typically, practitioners of that program tend to focus on a single language user, in isolation from both his social and physical environment. Jackendoff, as we saw, exemplified this approach; an approach that was strongly and effectively criticised by Putnam.

Linguistic behavior as a species of action

Bloomfield's naturalism was reductionist in nature. That is, he assumed that the linguist must assume the existence of a community of language users in a physical environment, but he held that the language users must be described using the vocabulary of behavioristic psychology and that the physical environment must be described using the vocabulary of the physical sciences. The behavioristic psychology that Bloomfield favored was intolerant of terms that smacked of "mentalism". Consequently, for Bloomfield, a speech act (such as his Jack and Jill example) must be described in the rather stilted and unfamiliar vocabulary of "stimulus" and "response" as we saw in section 2.2.

Advocates of the truth-conditional program disagree with Bloomfield's reductionist approach to the psychology of language users. They hold that language use must be considered a species of (largely) rational action. In order to understand this claim, we must understand what is meant by "action" as opposed to "mere behavior".

I follow David Shwayder and Donald Davidson in holding that both behaviors and actions are species of bodily movements, where the class of bodily movements includes the "null" movement.[193] Consider the following cases:

(i) Blair falls out of bed.

(ii) Blair curls up in a ball (when he is being tickled).

[193] D. Shwayder, The Stratification of Behavior, (Routledge and Kegan Paul, 1965). D. Davidson, "Agency", in R. Binkley, R. Bronaugh, and A. Marras, eds., Agent, Action, and Reason, (University of Toronto Press, 1971).

(iii) Blair loads the dart gun.

Case (i) is a mere movement of Blair's body; it is neither a behavior nor an action (although his precarious position that led to the fall may have been due to behavior, associated with dreaming perhaps). We recognize case (ii) as an example of instinctive behavior; although Blair is not intentionally curling up, it is something that his body is doing: it is a physiological response to the tickling. That is the hallmark of behavior: it is a movement that is mediated by physiological mechanisms. Physiological participation is lacking in the flight from the bed to the floor, and this rules out case (i) as an example of behavior.

Case (iii) is an example of behavior, but it has an additional characteristic that not all behaviors have. It is a movement that can be (and is in sentence (iii)) described as intentional on Blair's part. A bodily movement is an action if, under some description, it can be seen to be intentional on the agent's part. A bodily movement A can be seen to be intentional on an agent's part - and therefore an action - if the agent has a reason for the movement. Davidson characterizes a reason as follows:

> R is a primary reason why an agent performed the action A under the description D only if R consists of a pro attitude of the agent towards actions with a certain property, and a belief of the agent that A, under the description D, has that property.[194]

---

[194] D. Davidson, "Actions, Reasons, and Causes", The Journal of Philosophy, 1963, 60. Reprinted in D. Davidson, Essays on Actions and Events, (Oxford University Press, 1980), 5.

Blair has a desire (a pro attitude) to shoot the dart gun, and because he believes it must be loaded prior to shooting, he also has the desire to have the dart gun loaded. He also believes certain movements that he can make, which he would describe as 'loading the gun', will cause the gun to be loaded. Thus he has a reason for his body to move in this gun-loading fashion, and thus his movement qualifies as an action.

In order that we may speak of actions we must be willing to attribute beliefs and desires to agents. Beliefs and desires are examples, perhaps the central cases, of what Bertrand Russell called "propositional attitudes".[195] A propositional attitude is a term that is, or at least appears to be, a two-place predicate that is psychological in nature and relates a person to a proposition. Examples are 'believes', 'desires', 'hopes', 'fears', etc., as they appear in the following examples:

(i) Jeff believes that he is an ace pilot.
(ii) Jeff desires that he play with the flight simulator.
(iii) Jeff hopes that the computer is free.
(iv) Jeff fears that the computer is being used.

These examples appear to relate a person to the proposition expressed by the subordinate sentence. There are a number of unsettling issues associated with propositional attitudes when understood in this fashion. First, the metaphysical status of propositions is murky. Quine has challenged the proponent of propositional attitudes to provide precise identity conditions for these entities, and has expressed scepticism that such identity conditions are forthcoming. Of course, if the metaphysical status of propositions is in doubt, then their empirical status is in even worse shape. How can an empirical psychologist base his theories on propositional attitude psychology if it is not clear how to empirically identify propositions? Finally, there are logical difficulties with propositional attitude ascriptions like (i) - (iv). In languages conforming to the deductive patterns of classic first order predicate logic, Leibniz's Law prevails. This "law" holds that the substitution of a co-referring term in a sentence will produce another sentence with the same truth value as the original. Thus, substituting '3^2' for '9' in sentences of arithmetic will take truths into truths and falsehoods into falsehoods. However, example (ii) above is true, but

---

[195] B. Russell, An Inquiry into Meaning and Truth, (1940, reprint Penguin University Books, 1973), 159-160.

(v) Jeff desires that he play with the software package copyrighted by Bruce Artwick and Microsoft Corporation in 1988.

is false, even though 'The flight simulator (that Jeff wishes to play with) is the software package copyrighted by Bruce Artwick and Microsoft Corporation in 1988' is true.[196] Leibniz's Law does not hold for propositional attitude ascriptions, and consequently the well understood deductive patterns of first order logic do not apply to these types of sentences.

---

[196] There is one sense in which sentence (v) is true, i.e., the sense in which there is a program copyrighted by etc., such that Jeff wants to play with that program, even though he knows nothing about the ownership of the program. However, there is another sense in which (v) is clearly false, for Jeff himself would either deny its truth, or at a minimum he would request more information before assenting or dissenting. For a discussion of the two readings, see W. V. O. Quine, "Quantifiers and Propositional Attitudes", The Journal of Philosophy, 1963, 53, 177-187. However, for recent doubts about the validity of distinguishing the two readings, see S. Stitch, From Folk Psychology to Cognitive Science, (The M.I.T. Press, 1984).

On the face of it, then, propositional attitudes look like a poor bet as a basis for a scientific psychology. Bloomfield felt that linguistics must have a close relation with psychology, but it must be a scientific psychology that we turn to. Bloomfield's writing was always at its most passionate and invective when he was criticizing vague "mentalistic" formulations of other theorists.

But like it or not, it appears that the kind of psychology that we must use to characterize people as language users is propositional attitude psychology. That is, the use of language can be construed as a kind of action, but it cannot be construed as a kind of behavior. Of course, if action can be in some way "reduced" to behavior, then language use can be construed as a kind of behavior, but only, I submit, by first characterizing language use as action and then performing the reduction. In any case, in the next chapter I will argue that action cannot be reduced to behavior.

To see why language use cannot be directly construed as a species of behavior, let us consider a question addressed by H. P. Grice in his paper "Meaning".[197] Grice asked, in effect, what distinguishes social interactions that involve symbolic

---

chapter 6; and D. Dennett, "Beyond Belief", in A. Woodfield, ed., Thought and Object, (Oxford University Press, 1982).

[197] H. P. Grice, "Meaning", Philosophical Review, 1957, 66, 377-388. Reprinted in J. F. Rosenberg and C. Travis, eds., Readings in the Philosophy of Language, (Prentice-Hall, 1971).

communication from other types of social interaction. Grice wants to know what it is for a person to make an utterance and to "mean" something by making the utterance, such as when a person utters 'Fire!' thereby meaning to warn others of a dangerous fire. Grice is not interested in what might be called "natural meaning". A person's shrieking in pain "means" that he has been injured, and others may learn that he is in pain by hearing the shrieking, but this is "communication" only in the "natural" sense that falling leaves "mean" it is autumn. Rather, Grice is interested in the conditions under which a person uses an utterance (where "utterances" are broadly construed as any kind of behavior, vocal or otherwise) as an arbitrary symbol to (non-naturally) mean something. In other words, Grice wants to replace the ellipses in the following sentence with a set of conditions that will distinguish symbolic communication from other types of social interaction:

> By uttering x, person U non-naturally meant something
> if and only if ...

Bloomfield's response to Grice's problem can be deduced from his discussion of the Jack and Jill example, and would go something like this:

> By uttering x, person U non-naturally meant something if and only if U is experiencing a particular pattern of stimuli and responses, and through a process of conditioning U has developed a propensity to utter x when experiencing this pattern.

Bloomfield's analysans (i.e., what follows the 'if and only if') does not ascribe any propositional attitudes to the person; instead it uses the ascetic language of behavioral psychology. However this sort of analysis clearly will not do, for it fails to distinguish symbolic communication from almost any other type of behavior, let alone other types of social interaction. The analysans is simply a (crude) characterization of learned behavior, as it is viewed by behavioral psychology.

Other advocates of the behaviorist program in psychology such as Osgood, Mowrer, and Skinner have attempted to do better than Bloomfield and have tried to distinguish more carefully symbolic from non-symbolic behavior using only the terminology of the behaviorist program. These attempts have been strongly criticized by Chomsky, Fodor and others.[198] I will not review these attempts and criticisms, and simply note that I am in agreement with most of what Chomsky and Fodor have to say. Rather, I will point out the futility of these approaches by briefly examining how Grice proposes that we answer his question.

Grice proposes that symbolic interaction (cases of non-natural meaning) be characterized in terms of propositional attitude psychology. His suggestion is that:

(G) By uttering x, person U non-naturally meant something if and only if U intended the utterance of x to produce some (particular) effect in an audience by means of recognition of this intention.

Grice's analysans will not be particularly meaningful to the uninitiated. In order to understand it, let us first try to get away with something simpler. Assume that we are asked to provide an analysans of Grice's analysandum, and that we are free to use propositional attitudes. Most of us, in our pre-Gricean innocence, would suggest something like the following:

---

[198] See C. E. Osgood, Method and Theory in Experimental Psychology, (Oxford University Press, 1953); O. H. Mowrer, "The Psychologist Looks at Language", American Psychologist, 1954, 9, 660-694; and B. F. Skinner, Verbal Behavior, (Appleton-Century Crofts, 1957). For criticisms, see N. Chomsky, "Review of Verbal Behavior by B. F. Skinner", Language, 1959, 35, 26-58; and J. A. Fodor, "Could Meaning be an r_m?", The Journal of Verbal Learning and Verbal Behavior, 1965, 4, 73-81.

> By uttering x, person U non-naturally meant something if and only if U intended the utterance of x to produce some (particular) effect in an audience.

Thus, my uttering 'I have the measles'' counts as symbolic behavior because I intended, in uttering it, to produce some sort of effect in my doctor, viz., to have him come to believe that I have the measles. If I have red spots, and the doctor sees them, that may cause the doctor to come to believe that I have the measles, but the red spots do not count as an instance of symbolic communication because I did not intentionally have the spots for the purpose of making the doctor believe that I have the measles, as required by the analysans.

Still, as Grice points out, this analysis will not do. He notes that the analysans is satisfied by countless cases of social interaction that are not intuitively cases of symbolic communication. To use one of Grice's examples, if I leave A's handkerchief at the scene of a murder in order that the police should come to believe that A is the murderer, then I have satisfied the conditions specified above, but surely I did not, by leaving the handkerchief, mean that A is the murderer, nor does the handkerchief, or the leaving of it, constitute a symbol that A is the murderer.[199]

Grice suggests the following amendment to our preliminary analysis. In symbolic communication the utterer must not only intend to effect the audience in some way, the utterer must also intend that the audience recognize that the utterer is intending to so affect him. This additional qualification would rule out the case of "planting" the handkerchief as evidence. However, as Grice points out, even this amendment is not sufficient to rule out certain cases that do not involve symbolic communication. For suppose that I show you a photograph of your spouse compromising his or her marriage vows, and that in showing you the photograph I intend that you thereby come to believe that your spouse is unfaithful, and furthermore, I intend that you recognize that I did it intending that you form that belief. This case meets the amended conditions, but surely I did not, by showing you the photograph, mean that your spouse is unfaithful, nor does the photograph, or the showing of it, constitute a symbol of your spouse's infidelity.[200]

---

[199] H. P. Grice, op. cit., 440.

So Grice adds one further condition. In symbolic communication the utter must intend (1) to affect his audience in some way, (2) that the audience recognize the intention (1), and (3) that the recognition mentioned in (2) should be part of the audience's reason for being affected in the way mentioned in (1). (Analysis (G) is just a compact statement of this set of conditions.) To see how this third condition works, consider the photograph example again. Clearly, your recognition that I showed you the photograph because I wanted you to thereby come to believe that your spouse is unfaithful is not essential to your forming that belief; merely seeing the photograph left on the kitchen table would suffice for that. Consequently condition (3) rules out the photograph example. On the other hand, suppose that instead of showing you the photograph I had uttered the sentence 'Your spouse is cheating'. The sounds I produce will provide you with a reason to believe that your spouse is unfaithful, but only on the condition that you recognize that I produced those sounds with the intention of getting you to form that belief. (There will, of course, be other qualifications relating to my reliability as an informant, etc.) The same sounds produced accidentally by a squeaky bicycle pedal or by the wind rustling in the trees would not provide you with a good reason to come to that belief.

---

[200] Ibid., 440.

Grice does seem to be on to something here. His analysis of non-natural meaning seems to mark off precisely that type of communication in which symbols, rather than natural signs, are employed.[201]

I believe that Grice's analysis of non-natural meaning provides an adequate defense of the thesis that language use has to be, initially, at least, conceptualized as a species of action rather than a species of mere behavior. I accept that the complex propositional attitude (a self-referential intention) that Grice has identified is necessary in order to avoid the counterexamples discussed. Even if actions are reducible to behavior (which I doubt), it seems to me that behaviorists would have no chance of identifying, in behavioral terms, the delicate Gricean conditions required for symbolic communication unless they were first identified in propositional attitude terms. Consequently I agree entirely with the following slightly modified quotation from Brian Loar:

---

[201] It has subsequently been shown that Grice's analysis is subject to some rather subtle counterexamples, and complex alterations are required in order to save the analysis. A detailed account of these issues can be found in S. Schiffer, Meaning, (Oxford University Press, 1972). However, these revisions do not work against the main point I am trying to make, viz., that linguistic behavior has to be characterized in terms of propositional attitudes. Rather, the case is strengthened, because the required revisions are even more complex patterns of propositional attitudes.

> ...[Pragmatics] is a part of propositional-attitude psychology, and stands or falls with it. If propositional attitudes cannot be accommodated in a scientific conception of reality, then neither can semantics; but if they can, there is no need to cast about for anaemic approximations to our red-blooded intuitive [pragmatical] notions.[202]

In a nutshell, language use is a species of action and therefore the study of language use is committed to propositional attitude psychology (although it leaves open the question of whether propositional attitudes can in some way be "eliminated"). Note that this is not an especially unusual point. Many social sciences have a similar commitment. Sociology, economics and political science are, for the most part, based on the idea of a community of actors, each with their own "utility functions" and "subjective utilities". The work of the masters of social science - Adam Smith, Comte, Marx, Weber, Durkheim, Marshall, Freud, G. H. Mead, Keynes, Parsons, Neumann and Morgenstern - can be viewed as the continuous development and enhancement of a model of human society as composed of actors who make decisions and conduct themselves as a function of their beliefs and desires. The work of Grice (and of a good many other philosophers of language) shows that the study of natural language should take its place within this grand framework.

---

[202] B. Loar, "Two Theories of Meaning", in G. Evans and J McDowell, Truth and Meaning, (Oxford University Press, 1976), 138-139. The "slight modification" is that I have substituted the term 'pragmatics' for Loar's 'semantics'. The substitutions are indicated in square brackets. My reason for making this substitution will become clear in the subsection on "Linguistic Platonism".

Linguistic Platonism

Notwithstanding the previous point, languages should be understood as formal objects that are specifiable independently of their use in human societies. This is not a contradiction. The last point was that the use of language, i.e., the field of pragmatics, is a part of a social science that is based on the ascription of propositional attitudes. The present point is that the syntax and semantics of particular languages can, and should, be specified completely formally, i.e., independently of social science.

This point is perhaps made clearer by considering the phenomenon of the use of mathematical systems by human societies. Let us consider the use of arithmetic. Now some philosophers, e.g., John Stuart Mill, have argued that arithmetical truth is, at bottom, a matter of human psychology.[203] This view, known as "psychologism", holds that we know the truth of '2+2=4' on inductive grounds. All so-called deductive thinking is actually inductive, according to Mill.

---

[203] J. S. Mill, A System of Logic, (1843, reprint Hafner Publishing, 1950), Book II, chapter 6.

Mill's psychologism is not a popular position in the philosophy of mathematics. Opposed to it is "mathematical Platonism", as represented by Gottlob Frege. Frege warns against confusing mathematical truth, e.g., the fact that 2+2=4, with the history of the knowledge of that truth in various human societies. '2+2=4' was true, according to Frege, even before its truth was grasped by that first human (or protohuman) who stopped to reflect on arithmetical matters. '2+2=4' would be true even if humans had never existed, or even if it were physically impossible that there could be an organism capable of grasping arithmetical truths. According to Frege, confusion over mathematical facts and the psychological issue of how those facts are "grasped" is a fundamental mistake that must be avoided.

> Never take a description of the origin of an idea for a definition, or an account of the mental and physical conditions through which we become conscious of a proposition for a proof of it. A proposition may be thought, and again it may be true; never confuse these two things. We must remind ourselves, it seems, that a proposition no more ceases to be true when I cease to think of it than the sun ceases to exist when I shut my eyes.[204]

For Frege, the truth of '2+2=4' resides in the fact that numbers are formal objects (perhaps reducible to other formal objects such as classes, but nevertheless, formal objects) that have properties and relations to one another independently of any physical or psychological facts.[205]

---

[204] G. Frege, The Foundations of Arithmetic, trans. J. L. Austin, (Oxford University Press, 1950), introduction. For more on Frege's antipsychologism, see G. Frege, "The Thought: A Logical Inquiry", trans. A. M. Quinton and M. Quinton, Mind, 1956, 65, 289-311.

So we have two very different views of how to conceive of the phenomenon of the use of mathematical systems. Mill's "psychologism" holds that mathematical systems are, in fact, defined by their use. Since mathematical systems are themselves defined in terms of psychological processes, the use of those systems can obviously be defined in terms of nothing but ordinary psychological processes. Frege's "Platonism", on the other hand, holds that mathematical systems have to be defined independently of psychological processes. Consequently, the use of those systems cannot be described simply in terms of ordinary psychological processes. Rather, a human being can be said to be using arithmetic only if he (his psychological processes) are in some sort of relation to the formal system of arithmetic. For Mill, arithmetical practice is simply a property of a human being. For Frege, arithmetical practice is a relation between two things: a human being and the formal system of arithmetic. I believe that Frege is right.

It may be argued that Frege's psychology of arithmetic must be incorrect on the grounds that proper psychological theories should not assume the existence of entities other than subject of the psychological theory. As we saw in the last section, Putnam calls this the assumption of "methodological solipsism". My view is that one may hold this assumption if one chooses, but then one cannot have a psychology of arithmetic. Consider a ordinary four function electronic calculator. It would be possible to develop a "psychology" of the calculator that obeyed the principle of methodological solipsism; it would be a theory of how the calculator produces "responses", i.e., patterns on the display, as a result of "input", i.e., keypad presses. However, this theory, by itself, would not be a psychology of arithmetic for two reasons. First, the input-output function defined by the theory of calculator psychology could be interpreted as instantiating many other formal systems other than arithmetic. Secondly, the input-output psychology will not perfectly conform to arithmetic anyway. Four function calculators only conform to arithmetic within a certain range of numbers. Even within that range they are often constructed so that they give slightly incorrect answers.

---

[205] Although very few mathematicians and philosophers now support Mill's position, Frege's position is not universally accepted either. There is a minority dissenting opinion, called "intuitionism", which holds that the concept of mathematical truth cannot be divorced from the means by which that truth would be demonstrated. See, for example, L. Wittgenstein, Remarks on the Foundations of Mathematics, (1956, reprint The M.I.T. Press, 1967); and M. Dummett, "The Philosophical Basis of Intuitionistic Logic", Truth and Other Enigmas, (Duckworth, 1978). I do not wish to enter into this debate at this point, and I will simply assume that intuitionism is wrong.

If we really want to see the calculator as a device for performing arithmetical operations, then we have to see it as having a relation to arithmetic. Considered by itself it is not an arithmetic machine. Considered in the context of the two place relation

Instantiation (Calculator, Arithmetic)

it is an arithmetic machine. The relation of instantiation, by the way, must be understood as approximate instantiation. As mentioned above, any real life calculator will fail to perform arithmetical operations outside of a certain range, and even within the range, only approximately.

Similar considerations apply when one wishes to view human beings as users of arithmetic. One can insist on the principle of methodological solipsism if one chooses, but the resulting psychology will not be a psychology of arithmetic. If we want to view people as users of arithmetic, then we have to assume the Platonic doctrine that they are in some sort of relation to an independently characterized formal system. Furthermore, as Plato argued, the relation will be characterized by imperfection.

Returning to the issue of natural language, the Platonic claim is that a language is a formal object with a syntax and semantics (in Morris' sense of 'semantics': a relation between the items in the language and the world), and that some languages are used by populations of people. The use of a language (pragmatics) is a relation between the population and the formal language:

$$\text{Uses} (\text{Population}_x, \text{Language}_y)$$

The competing doctrine of psychologism holds that language use is simply a property of a population. No reference to an independently existing formal language is required:

$$\text{Language\_Use} (\text{Population}_x).$$

Note that Jackendoff's theory takes this approach to language. Bloomfield also views language use as a property of a population, and he to can be said to subscribe to psychologism, so long as radical behaviorism is viewed as a branch of psychology.

Advocates of the truth-conditional approach almost always adopt the position of linguistic Platonism. It should be noted, however, that other than the very general sort of remarks given above, it is not possible to give a really strong argument in advance for Platonism over psychologism. Jerrold Katz has attempted to argue for Platonism and against psychologism, but I doubt that these arguments would result in many conversions.[206] I think that the best argument for Platonism is that the resulting theory is superior to its alternatives. One demonstrates the point by developing the theory.

The essence of the Platonist view, and the main reason for adopting it, are neatly summarized in the following passage by David Lewis:

> I distinguish two topics: first, the description of possible languages or grammars as abstract semantic systems whereby the symbols are associated with aspects of the world; and second, the description of the psychological and sociological facts whereby a particular one of these abstract systems is the one used by a person or population. Only confusion comes of mixing these two topics.[207]

---

[206] J. J. Katz, "An Outline of Platonist Grammar", T. G. Bever, J. M. Carroll, and L. A. Miller, eds., Talking Minds, (The M.I.T. Press, 1984). Note that although Katz is a linguistic Platonist, he is not an advocate of the truth conditional approach.

[207] D. Lewis, "General Semantics", Synthése, 1970, 22, 18-69. Reprinted in D. Lewis, Philosophical Papers, Volume I, (Oxford University Press, 1983), 190.

Tarskian Truth Definitions as a Framework for Semantics

The fourth assumption is that the "truth definitions" that Tarski showed us how to construct for certain artificial languages are the appropriate framework for a Platonic theory of natural languages. That is, we should view languages as formal systems with the structure of a Tarskian truth definition. A particular population can be said to "use" one of these languages if certain psychological and sociological conditions are met.

Tarski's Constraints on a Definition of Truth

In order to introduce precision into his task, Tarski states that he views truth as a property of sentences (rather than beliefs, propositions, or anything else).[208] More precisely, he is interested in constructing a definition of the predicate 'is true', as it appears in constructions like:

'Nietzsche castigated Wagner' is true

Furthermore, the predicate 'is true' must be relativized to a language, so that we have

'Nietzsche castigated Wagner' is true-in-English

for it is possible that a sentence could occur in two languages, and it might be true in one language and false in another. Tarski's task, then, is to construct a definition of the predicate 'is true-in-L' where L is some particular language. In what follows I will sometimes drop the reference to a particular language, and speak simply of truth and the predicate 'is true'.

---

[208] A. Tarksi, "The Semantic Conception of Truth", Philosophy and Phenomenological Research, 1944, 4, 341-376, reprinted in L. Linsky, ed., Semantics and the Philosophy of Language, (University of Illinois Press, 1952), 14.

However, it should always be assumed that there is an implicit reference to some particular language.

Tarski's places two conditions on an adequate definition of truth. It must be (1) materially adequate, and (2) formally correct.[209]

(1) Material Adequacy

Let us use the term 'object language' to refer to the language for which we wish to construct a definition of truth. The term 'metalanguage' will refer to the language which we are using to formulate the truth definition. Tarski's condition of material adequacy is that a truth definition is adequate only if it implies, for every sentence of the object language, a metalanguage sentence of the form

(T) s is true iff[210] p

where 's' is replaced by a name of the object language sentence and 'p' is replaced by a metalanguage sentence that is the translation of the object language sentence. Thus, if we are trying to construct a truth definition of German in English, then it will be adequate only if it implies the English sentence:

'Schnee ist weiss' is true-in-German iff snow is white.

These sentences have come to be called "T-sentences" and Tarski's criterion of material adequacy has come to be called "Convention T". Note that T-sentences state necessary and sufficient conditions for the truth of their object-sentences.

---

[209] Ibid., 15-23.

[210] 'Iff' is short for 'if and only if'.

It is customary to say that T-sentences give the "truth-conditions" of their object-sentences.

It would be possible to meet Convention T simply by constructing a theory of truth that contains all the T-sentences as primitive axioms. The problem with this approach is that any language of theoretical interest has recursive syntactic rules, and therefore has an infinite number of sentences. If we want to construct a finitely stateable theory, then we will only be able to specify an infinite number of axioms by means of an axiom schema. However, (T) is not an axiom schema, since for any particular sentence s_i of the object language, the form (T) does not indicate what name we should use for s_i, nor does it indicate what the metalanguage translation of s_i should be. A true axiom schema, as used in the development of a theory of formal logic,[211] must be stated in such a way that there is no doubt about what sentences are and are not instances of the scheme. (T) does not provide that level of precison.

Consequently, in order to get a finitely stateable truth definition that meets Convention T, we must follow some other course. Tarski's idea is that in stating a truth definition we should follow the lead of syntax, in which the central tactic is to see sentences as made up of a finite list of sentential components (roughly, words) that are combined by recursive rules to yield an infinite number of sentences. Similarly, our truth definition must view the language as composed of a finite number of components each of which make a regular contribution to the truth-conditions of sentences in which they appear. And our truth definition must contain recursive rules that show how the truth-condition-determining properties of sentential components interact to determine the truth-conditions of whole sentences. The axioms of a truth definition will therefore consist of (i) individual axioms specifying the truth-condition-determining properities of sentential components, and (ii) axioms that specify how these components (recursively) combine to determine the truth conditions of sentences.

---

[211] As in A. G. Hamilton, Logic for Mathematicians, (Cambridge University Press, 1978), 71.

(b) Formal Correctness

Tarski also required that a truth definition be formally correct: that is, it should not lead to contradiction. However, this can happen very easily if we allow an object language to contain its own truth predicate. Assume that English contains its own truth predicate and we are trying to formulate a truth definition of English (object language) in English (metalanguage). Now consider the "self-referential" English sentence:

(A) Sentence (A) is not true.

The T-sentence for (A) will be:

> 'Sentence (A) is not true' is true iff sentence (A) is not true.

But since sentence (A) is identical with 'Sentence (A) is not true', it follows that:

> 'Sentence (A) is not true' is true iff 'sentence (A) is not true' is not true.

But this is a contradiction. This sort of "semantic paradox" will be generated in any language that contains its own truth predicate. The solution, according to Tarski, is to ban an object language's truth predicate from the object language itself, and to place it in the metalanguage. If one wishes to construct a truth definition for the metalanguage, then it will be necessary to use a meta-metalanguage that contains the metalanguage's truth predicate, and so on for an infinite hierarchy of languages.

Example: A Truth Definition for QL

Tarksi's method of defining a truth predicate will be demonstrated by specifying a truth definition for an artificial language QL. 'QL' stands for 'Quantificational Language', indicating that QL has the well understood logical structure of ordinary mathematical reasoning. The definition will be stated in English supplemented with some symbolization that will be explained on route, and some set theoretic notation. The purpose of the definition is to yield, for every sentence of QL, a T-sentence, i.e., a theorem of the form

s is true iff p

where 's' is a structural-descriptive name of a QL sentence, and 'p' is an English translation of the QL sentence named by 's'. The T-sentences are, of course, expressed in English.

The first step in the specification of QL is to provide a list of the symbols from which its sentences are constructed. They fall in the following five categories.

(1) A finite number of individual constants: 'a', 'b', ...

(2) A finite number of (one-place) predicates:

'F', 'G', ...

(3) A finite number of variables: 'x_1', 'x_2', ...

(4) Logical terms

(a) Sentential connectives: '~', '->'

(b) Quantifiers: 'V', 'E'

(5) Punctuation symbols: '(', ')'

It will prove convenient to provide names (in the metalanguage) for the various sets of symbols, so by stipulation 'CON' names the set of individual constants of QL, 'PRED' names the set of predicates, and 'VAR' names the set of variables. Furthermore, 'δ' is a metalinguistic variable which ranges over the members of CON, 'π' ranges over the members of PRED, and 'α' and 'β' range over the members of VAR. It will also be assumed that the members of VAR have been ordered, so that 'α_k' should be taken to refer to the k-th variable of QL.

As a means of facilitating comprehension of the definition English translations of QL symbols will be provided. The translation can be thought of as a function from QL symbols to English expressions, specifically, it is the following function:
{ <'a','Alfred Tarski'>, <'b','Richard Wagner'>, ...,
<'F','is a logician'>, <'G','is a musician'>, ..., <'x_1','x_1'>,
<'x_2','x_2'>, ..., <'~','it is not the case that'>,
<'->', 'if__then__'>, <'V','for every'>,
<'E','for at least one'>, <'(','('>, <')',')'> }.

The next task is to recursively define the class of sentences of QL. The strategy is as follows: first we define an *expression* of QL as any finite string of symbols of QL. (Note that the set of expressions will be denumerably infinite in size.) Next, infinite subset of expressions is recursively defined and this subset is called the set of *formulas* of QL. For convenience the metalanguage term 'FORM' will be used as a the name of the set of formulas, and the metalinguistic variables 'Φ' and 'θ' will range over the members of FORM. Next, an infinite subset of formulas is recursively characterized and this subset is called the set of *open formulas* of QL. The set of *sentences* is then defined as the set of formulas that are not open formulas. This strategy will be effected as follows: clauses (6) - (11) provide a recursive definition of the set of formulas, clauses (12) - (16) provide a recursive definition of the set of open formulas, and clause (17) defines the set of sentences.

(6) If $\pi \in$ PRED and $\delta \in$ CON, then $\ulcorner \pi\delta \urcorner \in$ FORM.

(7) If $\pi \in$ PRED and $\alpha \in$ VAR, then $\ulcorner \pi\alpha \urcorner \in$ FORM.

(8) If $\Phi \in$ FORM, then $\ulcorner {\sim}\Phi \urcorner \in$ FORM.

(9) If Φ, $\theta \in$ FORM, then $\ulcorner (\Phi\text{->}\theta) \urcorner \in$ FORM.

(10) If $\Phi \in$ FORM and $\alpha \in$ VAR, then $\ulcorner V\alpha\Phi \urcorner \in$ FORM.

(11) If $\Phi \in$ FORM and $\alpha \in$ VAR, then $\ulcorner E\alpha\Phi \urcorner \in$ FORM.

(In stating clauses (6) - (11) I have employed Quine's "quasi-quotes" (i.e., '$\ulcorner$', '$\urcorner$') as part of the metalanguage. These are to be thought of as a device which permits one to make claims using both metalinguistic variables and object language expressions. The governing stipulation is that '$\ulcorner y_0 \ldots . y_n \urcorner$' names an object language expression determined as follows: (i) If $y_0 \ldots . y_n$ are designatory expressions of the metalanguage, then the object language expression being named is the one obtained by writing the concatenation of the referents of $y_0, \ldots, y_n$. E.g., the expression '$\ulcorner \pi\alpha \urcorner$' names the object language expression formed by concatenating the object language predicate which is named by 'π' with the object language variable which is named by 'α'. (ii) If, on the other hand, any of $y_0, \ldots, y_n$ are object language symbols then the object language expression being named is the one obtained by concatenating the referents of $y_0', \ldots, y_n'$, where $y_i' = y_i$ if y_i is a metalanguage expression and $y_i' = 'y_i'$ if y_i is an object language expression. E.g., the expression '$\ulcorner {\sim}\Phi \urcorner$' names the object language expression formed by writing the object language symbol which is the referent of ''~'' followed by writing the object language expression which is the value of 'Φ'.)

The next step of the proposed strategy is to define the set of open formulas. These are formulas which have at least one free variable, the latter notion being characterized in clauses (12) - (16).

(12) If $\pi \in$ PRED and $\alpha \in$ VAR, then α is free in $\ulcorner \pi\delta \urcorner$.

(13) If $\Phi \in$ FORM and $\alpha \in$ VAR, then α is free in $\ulcorner {\sim}\Phi \urcorner$ iff α is free in Φ.

(14) If Φ, $\theta \in$ FORM and $\alpha \in$ VAR, then α is free in $\ulcorner (\Phi \text{->} \theta) \urcorner$ iff α is free in either Φ or θ.

(15) If α, $\beta \in$ VAR and $\Phi \in$ FORM, then α is free in $\ulcorner \mathrm{V}\beta\Phi \urcorner$ iff α is different than β and α is free in Φ.

(16) If α, $\beta \in$ VAR and $\Phi \in$ FORM, then α is free in $\ulcorner \mathrm{E}\beta\Phi \urcorner$ iff α is different than β and α is free in Φ.

This permits the following definition of the set of sentences.

(17) Any formula which contains no free variables is a sentence.

Clauses (1) - (17) define the syntax of QL, and they allow us to form a "structural-descriptive" name for every sentence of QL, i.e., these clauses allow us to formulate a distinct metalanguage expression for each sentence of QL, and one which reflects the syntactic structure of the QL sentence that it names. (Note, by the way, that the syntax of QL could easily be expressed as a set of "rewrite rules" in the manner of section 2.2. The style of syntax presentation used here conforms to the style usually used in textbooks of logic.[212] Because this style is a little more self-explanatory than rewrite rules, it is easier to see the close relation between syntactic rules and the semantic rules that will be specified below.)

We now specify the semantic rules that will provide a truth definition for QL. In order to get on with the truth definition it is necessary to give QL an "interpretation", i.e., a specification fo a referent for each of its constants, together with necessary and sufficient "satisfaction" conditions for each of its predicates. The intended interpretation for QL is given in clauses (18) and (19):

---

[212] These syntactic rules are modelled on those presented in D. Kalish, "Semantics", in P. Edwards, ed., The Encyclopedia of Philosophy, Volume 7, (Macmillan, 1967), 348-358. However, Kalish uses semantic rules that define truth-in-a-model, rather than truth. I have chosen to express the semantic rules without the notion of models, and have therefore adapted Kalish's semantic rules so that they rely on Tarski's notion of "sequences", rather than models. My formulation of the semantic rules is modelled on M. Platts, Ways of Meaning, (Routledge and Kegan Paul, 1979), 18-33.

(18) Individual constants

(i) 'a' refers to Alfred Tarski.

(ii) 'b' refers to Richard Wagner.

:

(19) Predicates

(i) An object μ satisfies 'F' iff μ is a logician.

(ii) An object μ satisfies 'G' iff μ is a musician.

:

Clauses (20) - (25) are a recursive definition of a semantic relation *satisfaction* which holds between sequences and *formulas* of QL. A sequence is an ordered set of objects with a denumerably infinite number of elements. A given object may be repeated any number of times in a sequence. It is convenient to stipulate that 's_k' refers to the k-th member of sequence S, and that 'S' $\approx_k$ S' means that sequence S' differs from sequence S in at most the k-th place. (Note that for every sequence S it will be the case that S $\approx_k$ S.)

(20) A sequence S satisfies $\ulcorner\pi\delta\urcorner$ iff the referent of δ satisfies π.

(21) A sequence S satisfies $\ulcorner\pi\alpha_k\urcorner$ iff s_k satisfies π.

(22) A sequence S satisfies $\ulcorner\sim\Phi\urcorner$ iff S does not satisfy Φ.

(23) A sequence S satisfies $\ulcorner(\Phi\text{->}\theta)\urcorner$ iff either S satisfies $\ulcorner\sim\Phi\urcorner$ or S satisfies θ.

(24) A sequence S satisfies $\ulcorner V\alpha_k\Phi\urcorner$ iff for every S' such that if S' $\approx_k$ S, then S' satisfies Φ.

(25) A sequence S satisfies $\ulcorner E\alpha_k\Phi\urcorner$ iff there is some S' such that S' $\approx_k$ S, and S' satisfies Φ.

Finally, clause (26) introduces the predicate 'is true in QL' and completes the definition.

(26) A sentence is true iff it is satisfied by all sequences.

Clauses (1) - (26) together constitute a truth definition for QL, i.e., a set of axioms form which it is possible to prove, for each sentence of QL, a theorem of the form

s is true iff p

where 's' is a structural descriptive name of the QL sentence and 'p' is a translation of s into the metalanguage, i.e., English (plus a little set theory).

Clauses (24) and (25) are perhaps a bit more obscure than the others. Consider clause (24) and its application to the QL sentence $\ulcorner Vx_1Fx_1\urcorner$. Clause (24) says that a sequence S will satisfy the sentence iff every sequence that differs from S in at most the first place satisfies $\ulcorner Fx_1\urcorner$. By clause (21) and (19.i), a sequence S' will satisfy $\ulcorner Fx_1\urcorner$ iff the first object in S' is a logician. Now this means that clause (24) states that sequence S satisfies $\ulcorner Vx_1Fx_1\urcorner$ iff the first object of every sequence that differs from S in at most the first place is a logician. But since every object will appear in the first position in one of these sequences, clause (24) states simply that a sequence S satisfies $\ulcorner Vx_1Fx_1\urcorner$ iff every object is a logician. (Note, by the way, that a given sequence S will satisfy $\ulcorner Vx_1Fx_1\urcorner$ iff every sequence satisfies $\ulcorner Vx_1Fx_1\urcorner$.) So clause (24), when combined with clause (26), is a precise way of saying that a universally quantified formula is true iff the formula within the scope of the quantifier ($\ulcorner Fx_1 \urcorner$ in our example) is satisfied by every object.[213]

To further enhance our understanding of the truth definition, let us now consider an example of the derivation of a particular T-sentence. Consider the QL sentence $\ulcorner Ex_2{\sim}Fx_2 \urcorner$, which, according to the translation function specified above, comes out in English as 'For at least one x_2, it is not the case that x_2 is a logician', which, in turn, may be considered as equivalent to 'It is not the case that every object is a logician'. Consequently, one of the T-sentences that has to be derivable from the truth definition of QL is the following:

Theorem: $\ulcorner Ex_2{\sim}Fx_2 \urcorner$ is true iff it is not the case that every object is a logician.

The derivation of this theorem goes as follows: By clause (2), $\ulcorner F \urcorner \in$ PRED, and by clause (3) $\ulcorner x_2 \urcorner \in$ VAR. By stipulation $\ulcorner x_2 \urcorner$ is the second member of VAR. Then by clause (7), $\ulcorner Fx_2 \urcorner \in$ FORM; by clause (8), $\ulcorner {\sim}Fx_2 \urcorner \in$ FORM; and by clause (10), $\ulcorner Ex_2{\sim}Fx_2 \urcorner \in$ FORM. By clause (15), $\ulcorner Ex_2{\sim}Fx_2 \urcorner$ contains no free variables, therefore by clause (26) we have:

(i) $\ulcorner Ex_2{\sim}Fx_2 \urcorner$ is true iff $\ulcorner Ex_2{\sim}Fx_2 \urcorner$ is satisfied by all sequences

---

[213] It might be wondered why clause (24) specifies the restriction on the set of sequences in the way that it does. Why not simply say that a sequence S satisfies $\ulcorner V\alpha_k\Phi \urcorner$ iff for every S', S' satisfies Φ? The restriction is required because a sentences with several quantified variables would not end up with the correct truth conditions without it. See D. Kalish, *op. cit.*, 352 for a lengthier explanation.

By clause (25) and the logical rule which allows the substitution of one side of a provable biconditional for another, (i) becomes:

(ii) $\ulcorner Ex_2{\sim}Fx_2 \urcorner$ is true iff for every sequence S there is some sequence S' such that S' $\approx_2$ S and S' satisfies $\ulcorner {\sim}Fx_2 \urcorner$

By clause (22) and the rule of substitution we get:

(iii) $\ulcorner Ex_2{\sim}Fx_2 \urcorner$ is true iff for every sequence S there is some sequence S' such that S' $\approx_2$ S and S' does not satisfy $\ulcorner Fx_2 \urcorner$

By clause (21) and the rule of substitution we get:

(iv) $\ulcorner Ex_2{\sim}Fx_2 \urcorner$ is true iff for every sequence S there is some sequence S' such that S' $\approx_2$ S and s_2' does not satisfy $\ulcorner F \urcorner$

Reflection on the nature of sequences indicates that 'for every sequence S there is some sequence S' such that S' $\approx_2$ S and s_2' ...' is equivalent to 'every object ...', so (iv) can be rewritten as:

(v) $\ulcorner Ex_2{\sim}Fx_2 \urcorner$ is true iff every object does not satisfy $\ulcorner F \urcorner$

Finally by clause (19.i) and the rule of substitution we get:

(vi) $\ulcorner Ex_2{\sim}Fx_2 \urcorner$ is true iff it is not the case that every object is a great logician

which is, of course, the theorem which we were trying to prove. This sample derivation of a single theorem of the truth definition does not, of course, prove that the theory will issue every theorem of the desired sort. To show that requires more advanced techniques. The purpose of the derivation of (vi) was to demonstrate the workings of the truth definition, not to provide a formal proof of its adequacy.

Truth Definitions for Natural Languages

The truth-conditional program for the study of natural languages assumes that Tarskian truth definitions can be given for natural languages. Thus, languages like English, German and Hopi are formal objects that happen to be used by particular human communities. Advocates of this program are making at least two very significant claims: first, that truth definitions can indeed be formulated for natural languages, in spite of the fact that natural languages are in many ways very different from the artificial languages that Tarski considered; and secondly, that all legitimate questions about natural language can be addressed within this program. The advocates of the program are claiming that we need only specify two things: a truth definition and the nature of the relationship between a language and the community that is using it. It will not be necessary to supplement these two things with, for example, a "theory of meaning" in order to address certain questions. Truth-conditions are enough, according to the advocates of this approach.

Why Truth?

Suppose one accepts that the proper starting points for the study of language are (1) a naturalistic attitude toward people and their environments, (2) the assumption that the use of language is essentially a species of human action rather than mere behavior - if language use can be successfully reduced to a behavioristic description it will only be because action in general can be so reduced, and (3) the assumption that natural languages should be viewed as abstract (Platonic) entities, and the users of languages are in a certain sort of relation to those abstract entities. Even if we do accept these three things, what reason do we have to accept the additional doctrine that Tarskian truth-definitions are the appropriate framework to describe the semantic characteristics of natural languages? It is not sufficient to merely state that "truth-conditions are enough". We need an argument to justify this move; to establish an essential connection between the concept of truth and the study of natural language.

Languages and Populations

David Lewis has provided an account of the relation between a population of people who use a language and the formal language that they use.[214] Lewis holds that the concept of truth is an essential element in this relation, and that once we understand the role of the concept of truth it becomes obvious that truth-conditional semantics is the appropriate framework for describing languages.

Lewis' account of the relation between languages and their users is based on his theory of social conventions, which is itself grounded in the model of the rational actor. To introduce that theory, suppose that A and B are carpenters building a house. In the course of their work they constantly have to measure lumber, and consequently they have to use some system of measurement. (In order to keep the example simple, assume that there are only two possible systems of measurement, metric and British.) Suppose further that each carpenter, if he were working alone, would be indifferent over which system to use, but since they are working together each prefers to use the same system that the other uses since this will facilitate their common enterprise. The situation is represented in the payoff matrix below.

---

[214] Lewis' theory of conventions, and linguistic conventions in particular, is presented in his Convention, (Harvard University Press, 1969) and in a revised form in "Languages and Language", Language, Mind, and Knowledge, Vol VII of Minnesota Studies in the Philosophy of Science, ed. by K. Gunderson, (University of Minnesota Press, 1975).

Figure 2.4.1
A Sample Coordination Problem

		A's Choice	
		Use British	Use Metric
B's Choice	Use British	5 5	1 1
	Use Metric	1 1	5 5

(Each cell represents the consequences to A and B of a pair of choices made by A and B. The upper right number represents A's payoff in that situation; the lower left number represents B's payoff.)

One way in which A and B can coordinate their actions to their mutual benefit is to make an explicit agreement to use one or the other system. However, an explicit agreement is not required provided that in the past both A and B have used only one system, say metric, and that this is common knowledge, i.e., each knows that the other has used only metric in the past, each knows that the other knows that he has used only metric in the past, and so on. This common knowledge gives each a good reason to continue using metric, and this is what they do, without the need of any explicit agreement to do so. Under such conditions Lewis would say that A and B conform to a convention to measure in metric, sustained by a common interest in more efficient carpentry.

More generally, Lewis analyzes conventions as follows (and here I am paraphrasing slightly):[215]

> A regularity R, in action or in action and belief, is a convention in a population P iff, within P, the following six conditions hold:
>
> (1) Everyone conforms to R.
>
> (2) Everyone believes that the others conform to R.
>
> (3) The belief that the ~~others~~ conform to R gives everyone a good and decisive reason to conform to R himself.
>
> (4) There is general preference for general conformity to R rather than slightly-less-than-general conformity - in particular, rather than conformity by all but any one. (This condition serves to distinguish cases of convention, in which there is a predominant coincidence of interest, from cases of deadlocked conflict.)
>
> (5) R is not the only possible regularity meeting the last two conditions.

---

[215] D. Lewis, "Languages and Language", in Language, Mind, and Knowledge, Vol VII of Minnesota Studies in the Philosophy of Science, ed. by K. Gunderson, (University of Minnesota Press, 1975), 5-6.

> (6) The various facts listed in conditions (1) to (5) are matters of common knowledge: they are known to everyone, it is known to everyone that they are known to everyone, and so on.

With this background, Lewis goes on to analyze the relation between speech communities and the languages that they use in the following manner:

> (A) language L is used by a population P if and only if there prevails in P a convention of truthfulness and trust in L, sustained by a common interest in communication.[216]
>
> To be truthful in L is to act in a certain way: to try to never to utter any sentences of L that are not true in L. Thus it is to avoid uttering any sentences of L unless one believes it to be true in L. To be trusting in L is to form beliefs in a certain way: to impute truthfulness to others, and thus to tend to respond to another's utterance of any sentence of L by coming to believe that the uttered sentence is true in L.[217]

The idea of this: Each of us would like to have a wide repertoire of communicative acts at our disposal (since enhanced communication brings many benefits). Our range of possible communicative acts is vastly increased if we consistently produce and interpret them as a function of the truth-conditions of the sentences of some language. Any language will do, so long as everyone conforms to the same language. Lewis' convention of truthfulness and trust is intended to characterized the conditions under which this conformity will be realized. Moreover, Lewis' theory has the virtue of not appealing to any mythical agreement by which the members of a speech community explicitly chose the language they use.

---

[216] Ibid., 10.

[217] Ibid., 7.

Donald Davidson has suggested that while it may be a contingent fact that Lewis' convention of truthfulness and trust in a particular language does hold in many language-using populations, these conventions are not necessary for linguistic communication.[218] What is essential for language use in a population, according to Davidson, is that language users be able to interpret each other; i.e., be able to assign meanings to each other's sentences, and for Davidson this means that they be able to assign truth conditions to their neighbor's sentences. But this could happen in a community of, say five people, where each spoke a different language - provided, of course that each person was capable of interpreting all five languages. Furthermore, suppose that each person was in the daily habit of introducing some small change in vocabulary or grammar into his spoken language. Provided that the changes were not too radical, the others could keep up to the changes by correlating them with aspects of the speaker's intentions and various environmental cues. In other words, it is conceivable that linguistic communication may take place in a community that does not manifest the regularities that are essential to Lewis' theory of conventions.

According to Davidson, the reason that we are likely to find Lewis' pattern of linguistic convention being manifested in an actual linguistic community is that

---

[218] D. Davidson, "Communication and Convention", in Enquiries into Truth and Interpretation, (Oxford University Press, 1984), 275-280.

> ... we do not have the time, patience, or opportunity to evolve a new theory of interpretation for each speaker, and what saves us is that from the moment someone unknown to us opens his mouth, we know an enormous amount about the sort of theory that will work for him - or we know we know no such theory. But if his first words are, as we say, English, we are justified in assuming he has been exposed to linguistic conditioning similar to ours...[219]

In other words, a pattern of Lewisian convention will often be observed because it makes interpretation simpler in a large community. However, as we have seen, linguistic communication is possible in communities that lack these conventions. What is essential is interpretation, not convention. (This is not to deny that Lewis' analysis is an insightful clarification of perhaps universal, albeit contingent, pattern.)

Lewis' analysis fails to provide us with the a link between Tarskian truth definitions and language use in human communities. However, I have said that Davidson believes that interpretation essentially involves assigning truth-conditions to sentences. If convention is not the link between truth definitions and language use, might interpretation be? And if so, why?

Truth and Interpretation

John McDowell writes:

> Understanding a language consists in the ability to know, when speakers produce utterances in it, what propositional acts, and with what contents, they are performing.[220]

---

[219] Ibid., 277.

[220] J. McDowell, "Bivalence and Verificationism", in Truth and Meaning, ed. by G. Evans and J. McDowell, (Oxford University Press, 1976), 45.

Thus if we understand German then we will have the ability to know, as we watch a performance of Die Walküre, that when Brünnhilde sings "Hunding fällt dich im Streit" she is warning Siegmund (propositional act) that Hunding will kill him today (content).[221]

According to theorists like McDowell, Davidson,[222] and Mark Platts,[223] the notion of linguistic meaning is a theoretical notion; in particular, if we have an adequate theory of understanding (or, to say the same thing, interpretation), then everything there is to be said about linguistic meaning will be contained within that theory. On this view, a theory of understanding/interpretation for German should allow the possessor of that theory to identify, from the sights and the sounds, that Brünnhilde is performing an act of warning, and that the warning has as its specific "content" that Hunding will kill Siegmund today. What these writers have argued is that such a theory requires a Tarskian truth definition as a central component. In the following paragraphs I shall rely on Platt's presentation of the argument (which is essentially a reconstruction of McDowell's.)

The first step is to recognize a point made by Frege: that every linguistic act has two components which we can call "force" and "content". The content of an utterance is the "strict and literal meaning" of the sentence(s) uttered. The force of an utterance is the "point" of the utterance; it is what one is "doing" when one is uttering something. The German sentence "Hunding fällt dich im Streit" has a fixed content or meaning, but depending on the situation, it can be uttered with many different purposes: to report, to warn, to commiserate, to deplore; and if the speaker has a more malevolent attitude, to mock, to defy, to challenge, and so on. Consequently, Platts says that an overall theory of understanding must have two components: a theory of force and a theory of content.[224]

---

[221] R. Wagner, The Ring of the Nibelung, tr. by A. Porter, (Faber and Faber, 1976), 119.

[222] D.Davidson, Inquiries into Truth and Interpretation, Oxford University Press, 1984), xiii.

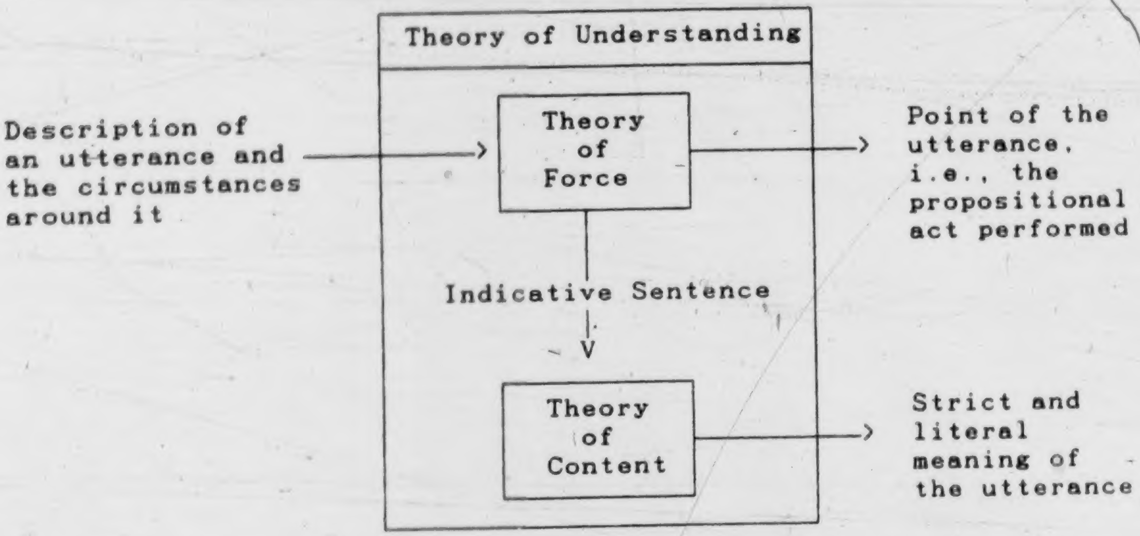
[223] M. Platts, Ways of Meaning, (Routledge & Kegan Paul, 1979), 58-63.

The overall theory takes specifications of utterances (and the circumstances around them), and yields specifications of the content and the point of those utterances. More precisely, the utterances and the circumstances around them are input to the theory of force. The theory of force yields a description of the utterance's force, i.e., a description of the act that the utterer was intending to perform. The theory of force also yields an indicative sentence, which is provided as input to the theory of content. The theory of content is used to determine the strict and literal meaning of the indicative sentence. This overall architecture is depicted in the diagram below.

---

[224] The theory of force has been elaborated in a body of literature known as "speech act theory". Two classic works are J. L. Austin, How to Do Things With Words, (Harvard University Press, 1962), and J. R. Searle, Speech Acts, (Cambridge University Press, 1969). A more recent work in speech act theory that explicitly endorses the truth-conditional approach discussed here is D. Holdcroft, Words and Deeds, (Oxford University Press, 1978). Holdcroft differs from Platts on one point: the principle of "semantic monism" that is discussed below. However, the difference is relatively minor. Holdcroft endorses the primacy of truth-conditions, and accepts the major theme of the argument to be given below.

Figure 2.4.2
Platt's Architecture for a Theory of Understanding


Theory of Understanding
Description of an utterance and the circumstances around it
Theory of Force
Point of the utterance, i.e., the propositional act performed
Indicative Sentence
Theory of Content
Strict and literal meaning of the utterance

Platts claims that if we assume this general architecture for a theory of understanding, plus a few additional assumptions, we will arrive at the conclusion that we require a Tarskian truth definition as the basis of the theory of content.

> We do not, so to speak, *start* from the bald doctrine that the meaning of a sentence can be specified by stating its truth-conditions; rather we *begin* from more general considerations... [and] we subsequently *discover* the connection with truth.[225]

Rather than simply assuming that a Tarskian approach is warranted, the claim is that it is a consequence of these more general considerations about the shape of a theory of understanding.

---

[225] M. Platts, *op. cit.*, 58.

But before examining how this consequence follows, more needs to be said about the role of indicative sentences in this scheme. The sentences of natural languages have various grammatical moods: i.e., the indicative, the imperative, the interrogative, etc. Platts claims that every sentence of a natural language, no matter what its mood, can be associated with an indicative sentence. Thus the interrogative 'Are you mine?' and the imperative 'Be mine' can both be associated with the indicative 'You are mine'. In requiring that non-indicative sentences first be converted into indicatives (which are, in turn, the subject matter of the theory of content), Platts is adhering to what he calls "the condition of semantic monism", which is the condition that

> ... the rules that determine the meaning of each and every sentence of in the language should all be of the same general kind. When we give an explanation of the meaning of any word or sentence, whatever the surface characteristics of that sentence (e.g. be it indicative, imperative, or interrogative), the explanation of that meaning always has to be of the same general type, has always to refer to the same kinds of considerations about that sentence. In Michael Dummett's terminology, there must be some key concept in the theory of meaning, some notion that figures in the explanation of the meaning of any sentence.[226]

The reason that Platts has selected indicative sentences as the input to the theory of content is that

> ... the indicative has a syntactic, semantic, and communicative completeness that the other moods lack: the absence of tense in imperatives; the eccentricity

---

[226] Ibid., 50.

> of a language with questions but no means to answer them; the dependence of commands upon the idea of their being satisfied; the crucial role of the indicative in language acquisition. All these invite the thought that the common element required by semantic monism will be indicative.[227]

However, selection of the indicative as the basis for a theory of content does not yet lead us to the conclusion that the theory of content has to take the form of a Tarskian truth definition. All that has been established so far is that indicatives will be the "input" to the theory of content, and a specification of the content of these indicatives will be its "output". This requirement can be recast as follows: the theory of content must produce, for every sentence of the object language, a sentence of the form

s ... p

where 's' is replaced by a structural descriptive name of the object language sentence, and 'p' is replaced by a meta-language sentence that expresses the content of any communicative act that a speaker would be performing in using s. The ellipses indicate that something must be filled in in order to make this a properly formed sentence, rather than just a pairing of objects.

> But if we now consider the filling between s and p, we can see that any filling that meets this condition of acceptability can acceptably be replaced by 'is true if and only if'. For the effect of the disquotation device, the truth-predicate, is to produce a sentence which can be used to say the very same thing, to perform the same propositional acts, as could the original sentence s prior to quotation or designation. That the truth-predicate is so insertable is a discovery: the general ruminations about the role of a theory of meaning within an explanation of [action] can be appreciated before the adequacy of the

---

[227] Ibid., 60.

> truth-predicate is realized. That truth functions as a disquotation device ... is all we need to appreciate, and all we need to say, to see that a theory of [content] is a theory of truth.[228]

This argument establishes that the theory of content should generate, as theorems, a T-sentence for every sentence of the object language. Once this has been established we are free to avail ourselves of Tarskian techniques to construct our theory of content.

A brief summary of this argument is perhaps in order. We started with a desire to construct a theory of natural language. We decided that this should take place within a larger theoretical context; the context of social science. Our social science will be characterized by a naturalistic attitude toward people and their environments. It was also decided that the use of language is best viewed as a species of action, rather than mere behavior. Furthermore, it was decided that languages cannot be reduced to action; rather, languages are abstract objects, and that the use of language should be construed as a relation between these abstract objects and their users. A theory of a particular natural language, we decided, should take the form of a theory of understanding; i.e., a theory that would allow its possessor to know which communicative acts, and with what content, a language user is performing when making utterances in the language. Such a theory of understanding will have two components: a theory of force and a theory of content. The theory of force will determine, for each utterance that a user of the language might make, the type of communicative act that is being performed, and an indicative sentence. The theory of content will determine, for the indicative sentence, what its content is. The theory of force will do this by specifying, for each indicative sentence of the object language in question, a theorem of the form 's ... p', where 's' is replaced by a structural-descriptive name of the indicative sentence, and 'p' is replaced by a meta-language sentence that gives the content of s. The ellipses can be replaced by 'is true if and only if', since the truth predicate has a disquotational feature, i.e., to assert 's is true' has the same import as asserting the sentence named by 's'. Thus the 'p' in 's is true iff p' states conditions that must be met if the sentence named by 's' is true; in other words, p gives the content of the sentence named by 's'. Thus, the theory of content must meet Tarski's criterion of material adequacy on truth definitions, and therefore the theory of content should take the form of a Tarskian truth definition.

---

[228] Ibid., 61.

Some Key Issues for the Truth-Conditional Program

The truth-conditional program has been described and defended in broad outline. In what follows I will briefly consider a number of issues and difficulties that the program faces. This list is not meant to be exhaustive, and the discussions of the issues I do raise are meant only as preliminary introductions. In the two decades that the program has been followed, a huge literature has developed; with contributions by champions and critics of the program. The following remarks are meant as only a very brief and partial introduction to that literature.

1. Truth-Definitions as Empirical Theories

When Tarski formulated his method of method of constructing truth-definitions of artificial languages his objectives were, of course, very different from those of present-day advocates of truth-conditional semantics for natural languages. One key difference is that Tarski took himself to be defining the truth predicate for particular languages. In order to construct this definition, Tarski made use of an unanalyzed notion, that of translation. (Recall Tarski's criterion of material adequacy: A truth-definition must generate, for every sentence of the object language, a theorem of the form 's is true iff p', where 's' is replaced by a structural-descriptive name of the object language sentence, and 'p' is replaced by a metalanguage sentence that is the translation of the object language sentence.) However, this will not do for contemporary advocates of the truth-conditional program. The truth-conditional program makes the essential claims that the notion of linguistic "meaning" or "content" is explicated when we show how to construct a theory of content, and a theory of content for a language is nothing more than its Tarskian truth-definition. But the truth-conditional program is in trouble if Tarskian truth-definitions rely on an unanalyzed notion of translation, for the obvious explanation of the concept of translation is that it is a mapping between sentences and expressions of two languages such that content is preserved. It appears, then, that we are caught in a circle.

Davidson has proposed a solution to this problem. We must "invert" Tarski's view of the relations between translation (and content) and truth. Our contemporary goal is to use the mechanisms of a Tarskian truth-definition to explicate the notion of content. Therefore we must not assume that concept in the truth-definition. Rather, the concept of truth must be taken as a primitive, unanalyzed notion.

> Our outlook inverts Tarski's: we want to achieve an understanding of meaning or translation by assuming a prior grasp of the concept of truth. What we require, therefore, is a way of judging the acceptability of T-sentences that ... makes no use of the concepts of translation, meaning, or synonymy, but is such that acceptable T-sentences will in fact yield interpretations.[229]

The last sentence expresses a subtle but extremely important point. Before we can construct a axiomatized Tarskian truth-definition for a language, we need an independent characterization of the T-sentences that the truth-definition is supposed to generate as theorems; i.e., we need to know which T-sentences are true (in the metalanguage) in advance of constructing the truth-definition (for the object language). Since Tarski took himself to be defining the predicate 'true' as it appears in T-sentences, he could not claim, in advance of giving the definition, to understand the T-sentences in which this predicate appears. To get around this problem, Tarski used the notion of translation to judge which T-sentences are true. This trick is not available to the modern advocate of the truth-conditional program, since this would lead to a vicious circularity. An alternate criterion, suggested by Davidson, is to assume that we (theorists of the object language) already understand the concept of truth, and therefore we understand, in advance of developing the axiomatic Tarskian theory, what the T-sentences are saying, and therefore can determine whether they are true or not. The last sentence in the quote hints that judging whether a candidate T-sentence is true or not involves more than just understanding it, and indeed Davidson does believe that there is more involved. The T-sentences proposed for a particular natural language must be judged against the evidence of the linguistic acts that are performed in the community that uses that language. This issue, the issue of how one would verify that a particular set of T-sentences is empirically supported by the evidence of linguistic behavior in a particular speech community, is discussed in detail in Chapter Four.

---

[229] D. Davidson, "Radical Interpretation", Dialectica, 1973, 27, 313-328. Reprinted in D. Davidson, Inquiries into Truth and Interpretation, (Oxford University Press, 1984), 150.

I will continue to use the term 'truth-definition' to describe the type of theory that contemporary advocates of the truth-conditional program are attempting to produce for natural languages. But keep in mind that they are not really in the business of defining truth. They assume a prior understanding of the predicate 'true' as it appears in T-sentences. Their real business is to come up with a axiomatized theory that generates all the true T-sentences, where a T-sentence is considered to be true if it is supported in the appropriate way by sociological evidence. (More on this in Chapter Four.)

2. Indexicality

The artificial languages that Tarski studied were free from indexical elements, but this is not the case with natural languages. Consider the sentence 'I like Wagner'. This sentence is true when uttered by some people but when uttered by others. Even for a single person it may be true when uttered at some points in time but not at other times. That is because the sentence contains three "indexical elements": the personal pronoun 'I', which has a reference that varies according to who is doing the uttering; the tensed predicate 'like', which is satisfied by different ordered pairs of things depending on when it is uttered; and the proper name 'Wagner', which may refer to the composer of Die Walküre, but it could refer to other individuals, depending on the context of utterance. If Tarskian truth-definitions are to serve as the basis of a theory of content for natural languages, then they will have to be adapted so they can accommodate indexical elements.

In an early paper, Davidson suggested a simple maneuver for accommodating indexical elements.[230] Rather than viewing the 'true' in T-sentences as a one-place predicate applying to object language sentences, we should view it as a three-place predicate applying to a sentence, a person, and a time. Thus, T-sentences should look like the following:

> 'I like Wagner' when (potentially) spoken by p at t is true iff p likes Wagner at t.

---

[230] D. Davidson, "Truth and Meaning", Synthése, 1967, 17, 304-323. Reprinted in D. Davidson, Inquiries into Truth and Interpretation, (Oxford University Press, 1984), 34.

Soon afterwards, David Lewis suggested expanding the range of "indices" beyond the two mentioned by Davidson: the speaker and the time.[231] He suggested adding places, audiences, things pointed at, segments of previous discourse, and other indices that require explanations that would take us too far afield.

Scott Weinstein has pointed out that the problem of indexicals is more complex that Davidson and Lewis have suggested, and that it cannot be solved by simply viewing truth as a multiple-place predicate (applying not only to a sentence, but also to a person, a time, and perhaps other things).[232] What Davidson's and Lewis' approach fails to take into account is that an utterance of a sentence may contain many occurrences of an indexical item, as in the sentence

That is a cat, and that is a dog, and that is a mouse.

The truth-conditions of such an utterance depend on what the individual 'that's refer to. This is not something that can be accounted for simply by treating truth as a multiple-place predicate as suggested by Davidson. In order to properly account for the truth-conditions of sentences containing multiply occurring indexical elements it is necessary to build the sensitivity to indexical reference right into the axioms of the truth-definition.

---

[231] D. Lewis, "General Semantics", in D. Davidson and G. Harman, eds., Semantics of Natural Language, (D. Reidel, 1972), 173-178.

[232] S. Weinstein, "Truth and Demonstratives", Nous, 1974, 8, 179-184.

Before we can do that, notes Weinstein, we must modify our criterion of material adequacy for truth definitions. Tarski's criterion (his "Convention T"), it will be recalled, is that an adequate truth-definition for a language L should imply a theorem, for every sentence of L, of the form

s is true iff p

where 's' is replaced by a structural-descriptive name of the object language sentence, and 'p' is replaced by a translation of s into the metalanguage.

We subsequently saw that Davidson's "inversion" of Tarski requires that the criterion be modified. Ignoring the issue of indexicals for now, according to Davidson a "truth-definition" (now considered as a theory of content rather than a definition of anything) is adequate only if it implies a theorem, for every sentence of L, of the form

s is true iff p

where 's' and 'p' are replaced as before, and where the theorem is both true and supported in the proper way by empirical evidence.

Weinstein suggests that a further revision of our criterion of material adequacy is required if we are to properly account for the phenomenon of multiple indexicality. The new criterion is that the truth-definition should imply, for each utterance of a sentence of the object language, a theorem of the form

> If u is an utterance of s and the referents of the indexical elements in u are $w_1, \ldots, w_n$ then u is true iff p

where u is replaced a structural-descriptive name of the utterance, and p is replaced by a metalanguage sentence derived from s by substituting for each of the i occurrences of the indexical elements in u and expression that refers to w_i, and such that the theorem is both true and supported in the proper way by empirical evidence.

Weinstein goes on to illustrate a formal method for constructing a truth-definition that meets this criterion of material adequacy, although we need not concern ourselves with the details in the present context. Weinstein also notes that the formal solution must be supplemented with a sociological-psychological theory (i.e., a pragmatical theory) of how the indexical elements contained in the utterances come to have the referents that they do. This adds a further requirement, and complication, to our overall theory of understanding.

As the work by Davidson, Lewis, and Weinstein shows, the phenomenon of indexicality can be accommodated by the truth-conditional program at the cost of additional theoretical complexity.

Before leaving this topic it should be noted that the phenomenon of indexicality is merely an additional complication in the lives the truth-conditional theorists. It also has a very positive side in that indexicality greatly facilitates the empirical testing of proposed truth-theories. As will be explained in Chapter Four, the key to testing empirical theories of truth is to study the conditions under which people assent to sentences. Even if we consider a wide variety of speakers in a variety of situations we will not find much variation of assent verdicts to a non-indexical sentence like 'Two plus eight is ten' (and where there is variation, it will probably set us down the wrong track), and therefore we will learn very little about the semantic role of words like 'eight' and 'ten' through a consideration of such sentences. However, if we study the assent patterns of indexical sentences like 'My brother is ten and I am eight' and correlating those patterns with certain features (i.e., the age) of the person who is assenting or dissenting, then we will start to learn facts about the semantic role of 'eight', 'ten', etc. Indexicality is therefore both a complication and a boon to the truth-conditional theorist.

3. Intensionality

Tarski developed his method of constructing truth-definitions for the extensional languages of mathematics. Following Cornman,[233] we can say that a sentence S is extensional if and only if:

> (1) The truth-value of a sentence which results from the replacement of any expression contained in the original sentence S by an extensionally equivalent expression (i.e., terms are extensionally equivalent if they refer to the same objects; predicates are extensionally equivalent if they are satisfied by the same set of objects) will not differ from that of the original sentence S under any conditions. (That is, the sentence S conforms to Leibniz's Law.)
>
> and
>
> (2) The truth-value of the sentence S, if it is compound or complex, i.e., if it contains coordinate main

---

[233] J. W. Cornman, "Intentionality and Intensionality", Philosophical Quarterly, 1950, 12, 44-52. Reprinted in A. Marras, ed., Intentionality, Mind and Language, (University of Illinois Press, 1972), 56.

> clauses or at least one subordinate main clause, is a function of the truth-values of the simple sentential elements which make up the compound or complex sentence. (That is, the sentence S conforms to the Principle of Truth Functionality.)

A sentence is intensional if it is not extensional. A language is extensional if all its sentences are extensional, otherwise it is intensional.

Frege pointed out that many sentences in natural language at least appear to be intensional. For example,

(i) Hegel believed that 9 > 7

is true, and

(ii) Hegel believed that the number of planets > 7

is false, even though

(iii) The number of planets = 9

is true. Sentence (i) appears to fail to meet condition (1), above, and therefore sentence (i) is an intensional sentence, and therefore it follows from our definition that English is an intensional language.

Tarski's original method of constructing a truth-definition was restricted to extensional languages, and it is not at all clear how one is to proceed in the face of seemingly intensional sentences like (i). Furthermore, (i) is not a freak occurrence. Natural languages are rife with such sentences. All sentences containing "propositional attitudes", such as 'believes', 'desires', 'hopes', 'fears', etc., appear to be intensional. Furthermore, sentences which express the alethic modalities (necessity and possibility), deontic sentences (i.e., sentences dealing with moral obligation), casual sentences, counter-factual conditionals, sentences about provability, and many other classes of non-psychological sentences appear to be intensional. How are we to proceed in the face of these sentences?

Frege's solution to the apparent phenomenon of intensionality was his famous theory of sense. Frege held that linguistic expressions such as names and predicates have references, but they also have senses. Thus '9' and the 'number of planets' have the same reference, but different senses. A sense of an expression give "the mode of presentation" of its reference, and

> ... the sense of an [expression] is grasped by everybody who is already sufficiently familiar with the language or totality of designations to which it belongs; but this serves to illuminate only a single aspect of the reference, supposing it to have one.[234]

Frege then claims that when a referential expression appears in the scope of a propositional attitude clause, that expression will refer to its customary sense. Thus sentence (i) and (iii) can be true and (ii) false without violating condition (1) because in sentence (iii) '9' and 'the number of planets' refer to their normal references, i.e., the number 9. But in (i) and (ii) these expressions appear in a propositional attitude context, and therefore refer to their customary senses. That is, in (i), '9' refers to the sense of '9' as it is customarily used, and in (ii) 'the number of planets' likewise refers to its normal sense. Since '9' has a different referent in (i) and in (iii),

---

[234] G. Frege "On Sense and Reference", Zietschrift für Philosophie and philosophische Kritik, 1892, 100, 25-50. Reprinted in P. Geach and M. Black, Translations from the Philosophical Writings of Gottlob Frege, (Basil Blackwell, 1977), 57-58.

(it can be thought of as a different word in the two contexts), Leibniz's Law does not apply to the trio of sentences, and therefore Leibniz's Law is not violated.

Frege's approach eliminates intensional sentences, but at the cost of introducing what are sometimes called "intensional entities", viz., senses. Quine, Nelson Goodman and others have long argued against the introduction of such entities into our ontology on the grounds of their lack of clarity. A Tarskian truth-definition for languages containing propositional attitudes can be constructed if we assume the existence of senses (in which case the axioms of our truth-definition will assign senses as the referents of terms, and senses as satisfiers of predicates), but how confident will we be that we understand the axioms, and in what sense can such a theory be considered empirical in nature? If it turns out that intensional entities are absolutely required in order to proceed with the semantic descriptions of natural languages, then we must adopt intensional entities in our analysis. But this should not be done lightly, and not without a serious examination of other alternatives.

Davidson has proposed another alternative. Following the lead of a number of theorists from Bertrand Russell to Noam Chomsky, he has argued that the surface form of natural language sentences is often misleading. Before investigating the semantic properties of a sentence it is first necessary to identify its underlying "logical form". If we do this properly, according to Davidson, we will see that the underlying logical form of the seemingly intensional sentences is actually extensional. There is no need to postulate intensional entities in the manner of Frege.[235]

As an example of Davidson's approach, consider his analysis of statements like

(iv) Galileo said that the earth moves.[236]

This appears to be just as intensional as (i) and (ii), for if 'earth' is replaced by certain coreferential terms, or if 'the earth moves' is replaced by certain logically equivalent sentences, (iv) will be transformed into a false sentence. Davidson says that the surface form is misleading, however. The underlying logical form is actually two sentences:

(v) The earth moves.
(Ex)(Galileo's utterance *x* and my last utterance
make us samesayers).

The second sentence is entirely extensional (provided that we allow utterances into our ontology and allow quantification over them), and therefore a Tarskian truth-definition can be provided quite easily. (The predicate 'samesaying' is a four place predicate that is satisfied by two utterers and an utterance made by each if and only if the two utterers said the same thing by their utterances.[237]) The central issue in such analyses is to defend the claim that (v) is a proper representation of (iv). Davidson does this by arguing that (v) can be abbreviated by definition to

---

[235] Actually Davidson does indulge in some ontological expansion. In his analyses of statements about human actions and singular causal statements he introduces events as a fundamental ontological category. See, for example, his "The Logical Form of Action Statements", in The Logic of Decision and Action, ed. by N. Rescher, (University of Pittsburgh Press, 1967).

[236] D. Davidson, "On Saying That", Synthése, 1968, 19, 130-146. Reprinted in D. Davidson, Inquiries into Truth and Interpretation, (Oxford University Press, 1984).

[237] Davidson says that the predicate is two-place, ibid., 106. However, I fail to see how the proper truth-conditions can

(vi) The earth moves.
Galileo said that.

He then uses an etymological argument to establish the equivalence of (vi) and (iv).

The Davidsonian approach to intensional contexts has a great appeal to those who are wary of intensional entities. But although progress has been made in some areas, many of the seemingly intensional sentences of English have still not been assigned a extensional logical form.

This lack of progress has lead some practioners of the truth-conditional program to adopt another approach to intensional contexts; an approach that is based on Saul Kripke's development of a Tarskian-like truth-definition for modal logic.[238] Kripke showed how to construct a Tarskian truth-definition for a language containing the intensional sentence operators 'necessity' and 'possibility'. Kripke uses the Leibnizian notion of a "possible world" as a key feature of his truth definition. Thus a sentence of the form ⌜Necessarily p⌝ is true iff p is true in every possible world; a sentence of the

---

be given unless all four objects are taken into account. In any case, this dispute is not particularly significant to my main point, which is to demonstrate how one can avoid the postulation of intensional entities by postulating an extensional logical form underlying the seemingly intensional surface form of sentences.

[238] S. Kripke, "Semantical Considerations on Modal Logic", Acta Philosophica Fennica, 1963, 16, 83-94. Reprinted in L. Linsky, ed., Reference and Modality, (Oxford University Press, 1971).

form ⌜Possibly p⌝ is true iff p is true in at least one world; and a sentence is true iff it is true in the actual world. Kripke's approach has been extended to a wide variety of intensional phenomena by a number of theorists, including Richard Montague, David Lewis, Robert Stalnaker, and many others.[239]

Possible-world semantics (i.e., truth-definitions that make reference to possible worlds) as an approach to the phenomenon of intensionality is very similar to Frege's theory of sense. Possible-world semantics makes short work of the logical problems associated with these sentences, but at the cost of introducing obscure entities (possible worlds), and additional problems that we could well do without (e.g., trans-world identification of individuals).[240] In fact, one possible world semanticist, David Lewis has argued that Fregean senses can be defined in terms of the constructs of possible-world semantics.[241]

I will not discuss the relative merits of Davidson's extensional approach versus the possible-worlds approach any further. It is sufficient to say that the intensionality of natural language poses difficult problems for the truth-conditional approach, and that there is considerable divergence of opinion as to how it should best be treated.

---

[239] See, for example, R. Montague, Formal Philosophy, ed. by R. Thomason, (Yale University Press, 1974); D. Lewis, Counterfactuals, (Harvard University Press, 1973); and R. Stalnaker, Inquiry, (The M.I.T. Press, 1987).

[240] M. J. Loux, ed., The Actual and the Possible, (Cornell University Press, 1979) is a collection of important papers discussing the ontological and metaphysical issues arising out of possible-world semantics.

[241] D. Lewis, "General Semantics", in D. Davidson and G. Harman, eds., Semantics of Natural Language, (D. Reidel, 1972). The idea of reconstructing Frege's theory in a formal system very similar to possible-world semantics was presented much earlier in R. Carnap, Meaning and Necessity, (University of Chicago Press, 1947).

4. Grammar

Our motivation for adopting the truth conditional approach is that it is a necessary component in a theory of understanding. A theory of understanding requires a theory of content, and, following Mark Platts, we argued that Tarskian T-sentences can be construed as matching sentences with their contents.

However, the requirement that a theory of understanding should include a theory of content demands no more than that we have a set of T-sentences available. A list of all the T-sentences for German, for example, would be sufficient to allow us to assign content to every indicative sentence in German.

But we went beyond giving a mere list; we suggested that we should follow Tarski and set down a finite set of axioms from which the T-sentences can be derived. Some of these axioms are syntactic in nature; i.e., they can be used to generate the sentences of the object language in question. Others are semantic in nature; mirroring the syntactic axioms, they establish relations between the expressions of the language and entities in the world, using the notions of reference, satisfaction and truth. Together, the set of syntactic and semantic rules can be called a "grammar"; i.e., to specify a Tarskian truth definition for a language is to specify its grammar. The question now to be asked is: why is it necessary, in constructing a theory of content, to go beyond a mere list of

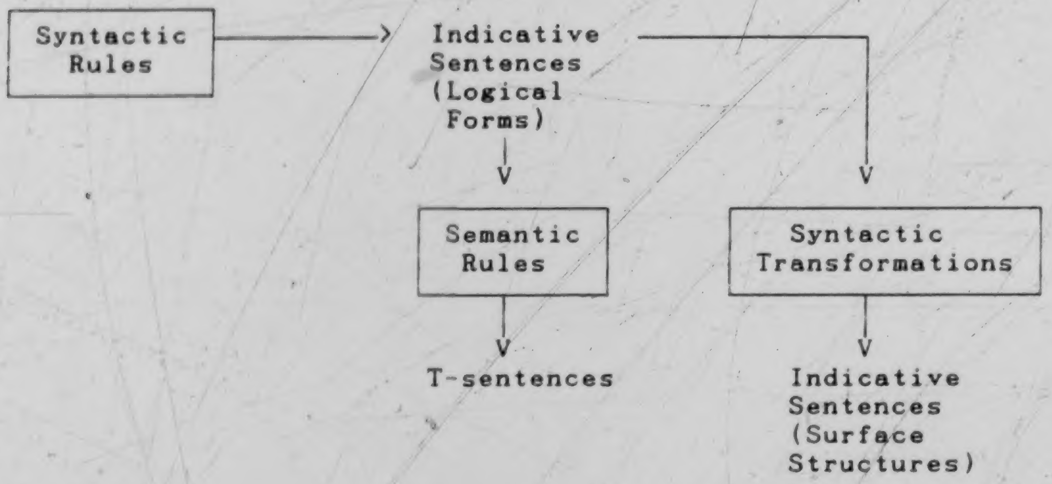
T-sentences; why is it necessary to provide a finite set of axioms that constitute its grammar.

There are two standard answers to these questions. One is that the number of sentences in a natural language is infinite, and therefore to account for the fact that people can learn natural languages it must be assumed that they learn a finite set of rules from which the infinite resources of language can be generated. The axioms of a Tarskian truth-definition can therefore be taken to represent the finite body of knowledge that a person must master if he is to learn a language. Let us call this the "learnability" argument. The other answer is that it is an essential feature of language that sentences are not wholly independent of each other; sentences appear to be composed of strings of elements (words) such that these elements reappear in various sentences, and they make a similar syntactic and semantic contribution to the various sentences in which they appear. This is a fact that should not go unnoticed in a theory of language, and it is therefore a virtue of the truth-conditional approach that it postulates axioms that govern the syntactic and semantic behavior of individual words. Let us call this the "compositionality" argument.

In the discussion of intensionality we saw that Davidson argued that it is inappropriate to try to construct a semantic theory to account for the "surface structure" of certain sentences. Certain sentences have an underlying "logical form" that may be somewhat hidden by the surface structure.

Consequently, the overall shape of a Davidsonian grammar will look something like this:

Figure 2.4.3
Structure of a Davidsonian Grammar


Syntactic Rules
Indicative Sentences (Logical Forms)
Semantic Rules
Syntactic Transformations
T-sentences
Indicative Sentences (Surface Structures)

This conception of grammar has obvious similarities to the work of transformational linguists, in that a distinction is being made between "surface" syntax and "underlying" syntax. Although Davidson does not formulate specific syntactic transformational rules in his papers, it is clear that a thorough formalization of his program will require the specification of such transformations, as will as the basic syntactic rules that generate the underlying logical forms.[242]

---

[242] For further discussion on the role of syntactic transformations within a truth-conditional framework see G. Harman, "Deep Structure as Logical Form", and D. Lewis, "General Semantics", both in D. Davidson and G. Harman, *Semantics of Natural Language*, (D. Reidel, 1972); B. H. Partee, "Some Transformational Extensions of Montague Grammar", *Journal of*

We see, then, that the overall structure of the theory of content developed by practioners of the truth-conditional program is not dissimilar from early work in transformational grammar.[243] However, there is a major difference between the way that transformational linguists and truth-conditionalists view their work. Most transformational linguists follow Chomsky in holding that grammatical theories are contributions to cognitive psychology, i.e., a formal grammar is construed as a theory of how the mind works. Truth-conditionalists, on the other hand, tend to avoid construing their grammars as having specific psychological implications.

The issue can be clarified somewhat by considering three different definitions of the term 'language'.

$\text{Language}_1 =_{\text{def}}$ A set of sentences

$\text{Language}_2 =_{\text{def}}$ A set of semantically interpreted sentences (specified by a list of T-sentences)

$\text{Language}_3 =_{\text{def}}$ A formal grammar that generates a set of semantically interpreted sentences (a Tarskian truth-definition)

Now transformational grammarians under the influence of Chomsky tend to see language use as primarily an psychological matter; i.e., according to the Chomskyans, individual people use languages, and communities of people use a language only in the derivative sense that the language is used by all the people in the community. Moreover, the Chomskyans hold that individual

---

Philosophical Logic, 1973, 2, 509-534; part one of W. G. Lycan, Logical Form in Natural Language, (The M.I.T. Press, 1986), and chapter six of M. Devitt and K. Sterelny, Language and Reality, (The M.I.T. Press, 1987).

[243] Compare figure 2.4.3 with the Chomsky's 1965 proposal for the organization of a grammar in figure 2.3.3.

people can use languages only because they possess, in some sense, an internalized grammar. Therefore there is a tendency to lean toward the language$_3$ concept in their work.

The truth-conditionalists, on the other hand, see language use as primarily a sociological matter; i.e., communities use languages, and individual people use a language only in the derivative sense that they belong to a speech community that uses that language. Moreover, when the truth-conditionalists say that a community uses a language they mean that "content" can be assigned to the linguistic actions of the community members in a particular way: that certain noises or marks have a standard "meaning", or more precisely, standard truth-conditions. Therefore, there is a tendency to lean toward the language$_2$ concept in their work. Of course the truth-conditionalists are extremely interested in constructing a grammar (a truth-definition) for particular languages, but the grammar is to be viewed as a tool of the theorist rather than the property of the language-user.[244]

---

[244] Readers familiar with Chomsky's distinction between linguistic "competence" and "performance" may feel that I am exaggerating the differences between the transformational linguists and the truth-conditionalists. For Chomsky argues that while psycholinguistic theories should strive to produce theories that account for linguistic "performance", and therefore their theories are descriptions of real psychological processes, the linguist, qua linguist, should not worry about psychological reality in a detailed manner. Instead, the grammars constructed by linguists are meant to describe an ideal speaker-hearer's "competence"; i.e., what he knows about language, not the detailed processes involved in speech production and perception. Chomsky writes "To avoid what has been a continuing misunderstanding, it is perhaps worth while to reiterate that a generative grammar is not a model for a speaker or a hearer. It attempts to characterize in the most neutral possible terms the knowledge of the language that provides the basis for actual use

In summary, many truth-conditionalists tend to view grammars instrumentally. That is, the output of the grammar - the T-sentences - have a sociological reality, in that the T-sentences will be supported by sociological evidence (as discussed in Chapter Four), and that they can be used to assign content to indicative sentences, and therefore they play a major role in a theory of understanding or interpretation. However, the grammar itself - the finite set of axioms that imply the T-sentences - does not have a psychological or sociological reality; rather it is a convenient tool of the theorist.

This instrumental attitude may seem to contradict one of the main motivations for constructing a finitely axiomatizable grammar in the first place; i.e., the argument that since languages are learnable, there must be a finite set of rules that are learned. However, the learnability argument does not imply that each of us must learn the clauses of some particular truth-definition in order that we may speak a language. Rather, it is simply that each of us must be characterized by some finitely realizable system that allows us to produce and perceive sentences. It does not follow that any two speakers of a language will share the same finite system, nor does it even follow that these internal systems can sensibly be described in the vocabulary of linguistic theory. It may turn out that a

---

of language by a speaker-hearer." N. Chomsky, Aspects of the Theory of Syntax, (The M.I.T. Press, 1965), 9. However, in spite of these claims, the transformationalist tradition has always had the tendency to assign psychological significance not only to the output of a grammar, but to the grammatical rules themselves. This contrasts sharply with most advocates of truth-conditional program, who generally assume no psychological significance whatsoever for the rules stated in their theories.

neurological vocabulary is required instead. However, even if all this is true, our goal of constructing a finitely axiomatizable grammar is still well-motivated, for the finite axiomatization will guarantee that the language is learnable, even if the axiomatization that the theorist produces is not the means by which real speakers do their learning.

2.5 CONCLUSION: WHORF'S THESIS AND THE STUDY OF LANGUAGE

We began this chapter with the observation that Whorf claimed that (1) each natural language has a unique set of syntactical categories, and (2) these syntactic categories have semantic correlates. In order to evaluate these claims, it was necessary to take a closer look at the notion of natural language and how it should best be studied. In particular, in order to make sense of Whorf's claims it is necessary to have a clear picture of how the semantics of natural language is to be conceived.

Whorf's own writings on the general concept of natural language are insufficiently clear to help us very much. Writing in the time when Bloomfield's ideas dominated American linguistics, Whorf seemed to accept the basic tenets of the Bloomfieldian program as far as syntax was concerned, but that it had somehow gone wrong on semantics. However, Whorf did not present his objections and alternatives very clearly.

In order to attain a better grasp of the concept of natural language we began, in section 2.2, by examining the linguistic program initiated by Leonard Bloomfield. Bloomfield's work can be aptly described as having a "hard-nosed scientific" character, as he was very reluctant to introduce linguistic concepts that could not be directly characterized in terms of empirical observations. His work, and contributions by a large number of linguists collectively known as "the Post-Bloomfieldians", resulted in the specification of "discovery procedures"; i.e., empirical operations on physical data that were intended to result in an objective description of the language used by a particular speech community. One characteristic of Bloomfield's approach was that it was assumed that detailed semantic descriptions of a language will have to be postponed until there are significant advances in the sciences that describe the objects to which natural languages refer.

I identified two objections to Bloomfield's program; one regarding semantics, and the other syntax. The problem with Bloomfield's semantics (assuming that the long wait for the appropriate advances in science is over) is that his semantics appears to take the form of a translation manual. That is, according to the Bloomfieldian program, we account for the semantic characteristics of a natural language, say English or Hopi, by translating it into the language of advanced scientific theories. However, the "translational" approach to semantics is unsatisfactory because it does not meet a minimal requirement of a semantic theory: that an adequate semantic theory should result in an understanding of the language that it is a theory of. I also noted that Whorf's attempts to "liberalize" Bloomfieldian semantics are too vague to take us anywhere.

The second criticism of the Bloomfieldian program is due to Chomsky. Chomsky and his disciples employed concepts of mathematical linguistics, and argued that the discovery procedures were limited to producing syntaxes of a certain complexity, but that natural languages are in fact characterized by syntaxes of a higher complexity. Although many of these arguments were not valid, they had a tremendous influence, so that during the sixties Chomsky's "transformational linguistics" became the dominant force in American linguistics.

In section 2.3 I examined how semantics is studied within the Chomskyan framework. However, in order to understand semantics as it is conceived within this framework it is necessary to develop some appreciation of Saussurian structuralism (which predates Bloomfield's work), for many of Saussure's assumptions have been incorporated into the Chomskyan framework.

Ray Jackendoff's recent semantic theory was examined as a representative example of this marrying of Saussure and Chomsky. It was argued that if we strip out some of the nativistic elements of Jackendoff's theory, we end up with a theory that is compatible with Whorf's Linguistic Relativity Hypothesis. I ventured the hypothesis that Whorf would have been happy with such a theoretical framework.

However, I criticized Jackendoff's theory as follows. As he formulates it, the theory is committed to an untenable Kantianism. But even when stripped of its Kantianism, it leads to the "traditional" theory that concepts are in the head, and that concepts are reference determiners. I used arguments developed by Putnam to show that this traditional theory is not tenable. This led to dissatisfaction with the program of "conceptual semantics", as described above.

Finally, in section 2.4, I turned to a more recent program: the truth-conditional program for the study of natural language. This program is based on the assumption that a theory of a natural language be incorporated within a larger theory of social action; that is, a theory that views people as engaging in actions and interactions as a function of their beliefs and desires. More particularly, a theory of natural language should be a theory of understanding (or interpretation), such that when a member of the community utters an expression of the language, the theory should tell us what (typically social) ends he was intending to accomplish by that utterance, and what content the utterance has. The latter requirement is met by a Tarskian truth-definition, construed not as a definition of the truth-predicate for that language, but as an empirical theory of the "content" of the language. Although the truth-conditional program faces a number of difficult problems, this program was endorsed as the correct approach to the study of natural languages.

Now, what is the relation between this program and Whorf's Linguistic Relativity Hypothesis? It was noted at the outset of this chapter, and again at the outset of this section, that Whorf held a realistic attitiude toward grammar; i.e, he held that languages have a unique set of syntactic categories, and these in turn have unique semantic correlates. However, in our discussion of grammar in the truth-conditional program we noted that within that program grammar tends to be construed instrumentally, not realistically. That is, the truth-conditionalist holds that the testable element of the theory is the set of T-sentences that the theory implies (we shall see how in Chapter Four), but the axioms of the theory are not themselves testable. Different axiom schemas (i.e., different grammars) are acceptable provided that they generate the same T-sentences.

Assuming the correctness of this view, it is clear that Whorf's Linguistic Relativity Hypothesis will have to be somewhat modified if it is to be considered within the context of the truth-conditional program. Instead of focussing on grammar, as Whorf does, we must focus on the consequences of the grammar, i.e., the T-sentences. If the Linguistic Relativity Hypothesis is true, and language A is very different from language B, then the set of T-sentences for A must, in some way, express very different things than language B. This shift from grammar to T-sentences will be taken up in more detail in Chapter Four.

The reader may feel that the switch in focus from grammar to T-sentences has been too quick. One may be influenced by Chomskyan arguments from language learnability to the conclusion that certain syntactic structures are innate, or one may be influenced by psycholinguistic research that suggests that certain parsing strategies are common to all users of a language. It may therefore be felt that a realistic account of grammar will be forthcoming from the psychology of language. The problem with these arguments is that even if they are sound (and this is certainly questionable at the present time) they apply only to syntax. If we want an account of semantics - and we do if we want to make sense of the Linguistic Relativity Hypothesis - then we must employ what appears to be the most reasonable semantic theory, and that, I have argued, is truth-conditional semantics. Therefore we need a grammar that gets the truth-conditions of sentences right. But there is no guarantee that such a grammar will be entirely consistent with the results of psycholinguistic research. For example, we have seen that in his account of the sentences containing the 'saying that' construction, Davidson postulates an underlying logical form that differs from the surface syntax of the sentence. However, it is conceivable that psycholinguistic research may discover that the psychological parsing process involved in perceiving that sentence is more compatible with the more obvious surface syntax. But in spite of that, we would still have to stick with the Davidsonian syntax in our grammar (assuming that we are entirely happy with Davidson's account of this fragment of English), for our objective is to get the truth-conditions right so that the grammar can take its place in a theory of understanding, which takes its place in turn in a theory of social action.

What we are contemplating is a potential tension between psycholinguistics and the theory of social action. Since the Linguistic Relativity Hypothesis essentially involves the idea of semantic content, it must be assessed in the theoretical context of a theory that accounts for content, and that must be (if the reasoning on the foregoing pages is sound) a theory of social action that incorporates a Tarskian theory. If the pronouncements of psycholinguistics suggest a syntactic theory that differs from the syntactic theory incorporated in our truth-definition, then so be it. We may have to accept that there will be a split between natural sciences of man (to which biology, and thus psycholinguistics, belongs) and the cultural studies, wherein only the latter can assign "content" to men, their actions, and their symbols. This theme will be developed further in the next two chapters.

THE UNIVERSITY OF ALBERTA

THE LINGUISTIC RELATIVITY HYPOTHESIS

by

DONALD H. MOTTERSHEAD ©

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

DEPARTMENT .....SOCIOLOGY...............

EDMONTON, ALBERTA

FALL, 1989

TABLE OF CONTENTS

PAGE

VOLUME II

CHAPTER THREE - THOUGHT **271**

3.1 THE GEISTESWISSENSCHAFTEN 271

3.2 NON-PROPOSITIONAL THEORIES OF THOUGHT 281

3.3 REDUCIBILITY, SUPERVENIENCE, AND PROJECTIBILITY . 295

3.4 FODOR'S FUNCTIONALISM 317

3.5 AUTONOMOUS SOCIAL SCIENCE 343

CHAPTER FOUR - RELATIVITY **361**

4.1 INTRODUCTION 361

4.2 ATTITUDES, MEANING AND PHYSICS 365

4.3 RADICAL INTERPRETATION 399

4.4 DAVIDSON ON CONCEPTUAL RELATIVITY 420

4.5 RELATIVITY OF REFERENCE 435

4.6 RELATIVITY OF REASONING 466

4.7 CONCLUSION 498

BIBLIOGRAPHY **506**

APPENDIX 1 - THE RESEARCH ON THE COLOR DOMAIN **522**

APPENDIX 2 - A SUMMARY OF THE ARGUMENT **561**

LIST OF TABLES

Table	Description	Page
A.1.1	Summary of the Color Domain Experiments	556

LIST OF FIGURES

Figure	Description	Page
3.3.1	Relation of the Geisteswissenschaften to the Natural Sciences	279
3.3.1	Two Sciences Explaining the Same Causal Relation	305
3.3.2	Two Sciences Explaining the Same Causal Relation According to the "Strong" Reading of Hume's Principle	307
3.3.3	Two Sciences Explaining the Same Causal Relation According to the "Weak" Reading of Hume's Principle	308
3.4.1	Fodor on the Relation Between Psychology and Physics	321
4.2.1	The Dual Method Alternative	392
4.2.2	The Single Method Alternative	393
4.2.3	The Single Method Alternative (Again)	397
A.1.1	Fishman's Systemization of the Whorfian Hypothesis	524

LIST OF SYMBOLS

Logical Implication	->
Logical Equivalence	≡
Logical Conjunction	&
Logical Negation	~
Logical Disjunction	v
Universal quantifier	(x)
Existential quantifier	(Ex)
Quine's quasi-quotes	⌜ ⌝
Set brackets	{ }
Ordered sets	< >
Element of a set	∈
Null set	∅
Set Union	U
Set Intersection	∩
Syntactic Production Rule	->
Syntactic Derivation	=>, =>>

CHAPTER THREE - THOUGHT

3.1 THE GEISTESWISSENSCHAFTEN

The previous chapter addressed the topic of natural language, and in particular, how natural language should be conceptualized and studied. I argued in Chapter Two that truth-conditional program is the most promising paradigm for the study of natural language. One of the fundamental features of that program is that it must be viewed as a component of a larger theoretical enterprise: a general theory of social action.

Natural language is, of course, the "independent variable" in Whorf's Linguistic Relativity Hypothesis. In this chapter our attention turns to the "dependent variable": thought. Given the results of the last chapter, my starting point for this chapter is that the account of "thought" that we come up with must also be integrated into the same larger theoretical enterprise, a general theory of action.

To begin, let us consider some general issues about this theoretical enterprise. Social science - in the broad sense that includes sociology, economics, history, political science, social psychology, cultural anthropology, comparative religion, etc. - has a long history that goes back to Plato. However, it was not until the latter part of the nineteenth century that a serious discussion emerged, primarily in Germany, regarding the epistemological basis of the social sciences. Julien Freund identifies the central issue in this discussion as follows:

> The bone of contention was the status of the human sciences (also called the historical sciences, the social sciences, the intellectual sciences, the cultural sciences, etc.): should they, as the positivists claimed, be assimilated with the natural sciences or, on the contrary, be regarded as wholly autonomous?[245]

In struggling with this question a number of theorists, including Wilhelm Windelband, Heinrich Rickert, Wilhelm Dilthey, Karl Jaspers, and Max Weber, began to identify the subject matter of the human sciences with the possession of "subjective meaning"; i.e., human behavior or artifacts fall within the scope of the human sciences if they are somehow endowed with "subjective meaning". The methodological activity of determining subjective meanings was called Verstehen, which is translated as 'understanding' or 'interpretation'. The Naturwissenschaften (i.e., the natural sciences) are focussed solely on objective description and explanation, whereas the Geisteswissenschaften (i.e., the social sciences) must also incorporate a method of "understanding" in order to explicate "subjective meanings", or so the story goes.

---

[245] J. Freund, *The Sociology of Max Weber*, (1968, reprint Vintage Books, 1969), 37-38. Note, by the way, that two historically important terms should be added to Freund's list of synonyms. John Stuart Mill coined the term 'the moral sciences' to refer to the social sciences in Book VI of his *A System of Logic*, (1843). Around the middle of the nineteenth century Wilhelm Dilthey introduced the term *Geisteswissenschaften* as a translation of Mill's term. The essence of the debate in late nineteenth century Germany was the relation between the *Geisteswissenschaften* and the *Naturwissenschaften*, or natural sciences. For an excellent short introduction, see H. P. Rickman, "*Geisteswissenschaften*", in P. Edwards, ed. *The Encyclopedia of Philosophy*, (Macmillan Publishing, 1968), Volume 7, 275-279. Also see H. Holbern, "Wilhelm Dilthey and the Critique of Historical Reason", in W. W. Wagar, ed., *European Intellectual History Since Darwin and Marx*, (Harper and Row, 1966).

Max Weber wrote of these issues, but he did not tarry long over the epistemological foundations of the social sciences. Assuming the basic validity of the idea of subjective meaning and the method of Verstehen, he began to develop a system of sociological concepts based on the core concept of subjectively meaningful action.[246] He then used these concepts as the framework for his monumental studies in history, sociology and economics. Given the universally acknowledged importance of these studies, his conceptual framework came into common use in the social sciences.

In the twentieth century the American sociologist Talcott Parsons was responsible for a major review of the conceptual underpinnings of the social sciences, from a basically Weberian perspective. In his The Structure of Social Action, Parsons undertook a historical study of the development of the "action" framework, culminating in a detailed study of Weber's contribution.[247] Later, Parsons, together with his collaborator, Edward Shils, published his own system of basic concepts, a system that incorporated a more detailed psychological model than what Weber had proposed.[248]

---

[246] M. Weber, Basic Concepts in Sociology, tr. by H. P. Secher, (The Citadel Press, 1962). This is a translation of chapter 1 of Weber's Wirtschaft und Gesellschaft, (1925).

My view is that Weber's conceptual system is of enormous value to the practising sociologist or historian, but that it is somewhat cloudy when it comes to addressing the question of the epistemological status of the Geisteswissenschaften. Parson's reconstruction does not help clarify things (and in fact, because of an unnecessary tendency toward complication, he makes things more difficult for the practising social scientist). In order to get a better grasp on the epistemological questions I believe that we should turn to a line of research initiated by Franz Brentano and Gottlob Frege.

As outlined in section 1.4 of this work, Brentano revived the Medieval notion of "intentionality", which he believed is required to characterize the essential nature of mental phenomena. He argued that what distinguishes mental phenomena from physical phenomena is that the former, but not the latter, are always directed toward some content; they are "about" that content. Thus wanting a holiday in San Diego is an example of a mental phenomenon in virtue of the fact that a holiday in San Diego is what the wanting is "about".

---

[247] T. Parsons, The Structure of Social Action, 2 Volumes, (1937, reprint The Free Press, 1968).

[248] T. Parsons and E. A. Shils, "Values, Motives, and Systems of Action", in T. Parsons and E. A. Shils, eds., Toward a General Theory of Action, (Harvard University Press, 1951).

Frege refined this thesis somewhat. He claimed that our central mental states (he called them 'ideas') are those that have "thoughts" as their objects. For example, believing that snow is white is an "idea", according to Frege, and the thought that snow is white is the "content" of the idea. What are thoughts? They are characterized in the following (very important) passage:

> Without wishing to give a definition, I call a thought something for which the question of truth arises. So I ascribe what is false to a thought just as much as what is true. So I can say: the thought is the sense of the sentence without wishing to say as well that the sense of every sentence is a thought. The thought, in itself immaterial, clothes itself in the material garment of a sentence and thereby becomes comprehensible to us. We say that a sentence expresses a thought.[249]

Frege's claim is that our central mental states are those that have, as their objects, something that has a truth value. Sentences have truth values - however, Frege would insist that they have truth values only because they have meanings (or "senses") that determine truth values. So Frege is claiming that our fundamental mental states are those that have, as their object, something that can be expressed in a sentence. It is common to use the term 'proposition' to describe the language-independent content of a sentence. So Frege's claim can be summed up as the claim that insofar as our mental states have content in Brentano's sense, that content is propositional in nature. That is, our fundamental mental states are the propositional attitudes, e.g.,

---

[249] G. Frege, "The Thought: A Logical Inquiry", tr. by A. M. and M. Quinton, Mind, 1956, 65, 289-311. Reprinted in P. F. Strawson, ed., Philosophical Logic, (Oxford University Press, 1967), 20.

Bernie believes that the social sciences are autonomous,
Gwynn grants that the social sciences are autonomous,
Leo laments that the social sciences are autonomous,
Richard realizes that the social sciences are autonomous,
etc.,

where a propositional attitude, as expressed by such "attitudinal" terms as 'believes', 'grants', 'laments', etc., is a state that puts the bearer of the mental state into some sort of relation to a proposition, e.g., the proposition expressed by the sentence 'the social sciences are autonomous'. (This should be construed as only a first rough cut at what a propositional attitude is. We already saw in the discussion of intentional sentences in Chapter Two that adding propositions to our ontology is not acceptable to most of us, and therefore we may try some other tack, such as the one that Davidson used in analyzing sentences containing 'said that'. However, any such analysis must conform to the "spirit" of the claim that a propositional attitude is a relation between a person and a truth-value bearing proposition.)

Propositional attitude ascriptions are, I believe, all that we need to characterize the "subjectively meaningful action" that is at the base of Weber's system of concepts. In other words, propositional attitude ascriptions are the basis of all the concepts of sociology, economics, political science, etc., can be constructed. (Decision theory, which is theory about how actions are determined as a function of propositional attitudes, will play a big role here.) Now that we realize, thanks to Brentano and Frege, that the Geisteswissenschaften are based on propositional attitude ascriptions, we can return to the question that so engaged late nineteenth century German intellectuals: the relation of the Geisteswissenschaften to the Naturwissenschaften.

More to the point, we can relate the nineteenth century debate in Germany to recent work by analytical philosophers. Over the past three or four decades there has been a tremendous amount of effort invested by philosophers interested in determining how the propositional attitudes fit into our world view, and in particular, how they relate to our scientific world view. Opinion varies on this issue. There are three main theories about how the propositional attitudes (and thus the Geisteswissenschaften) might fit into our scientific world view. First, the attitudes might be somehow "explainable" in the natural sciences. For example, the attitudes might reducible to a "behavioral biology" in a manner similar to the reduction of thermodynamics to statistical mechanics.[250] On the other hand, if the attitudes are not reducible in this manner or "explainable" in some other way (and assuming that we are materialists and that we have a healthy respect for the natural sciences), then we seem to have two further options. We may decide to retain the attitudes and develop the

---

[250] E. Nagel, The Structure of Science, (Harcourt, Brace and World, 1961), 338-345.

Geisteswissenschaften as an activity separate from the Naturwissenschaften. Or, alternatively, we may decide to give up on the Geisteswissenschaften.

These alternatives were laid out very clearly in 1960 by Quine in following passage. (Quine's scepticism about the possibility "reducing" propositional attitudes to "hard science" is quite apparent.)

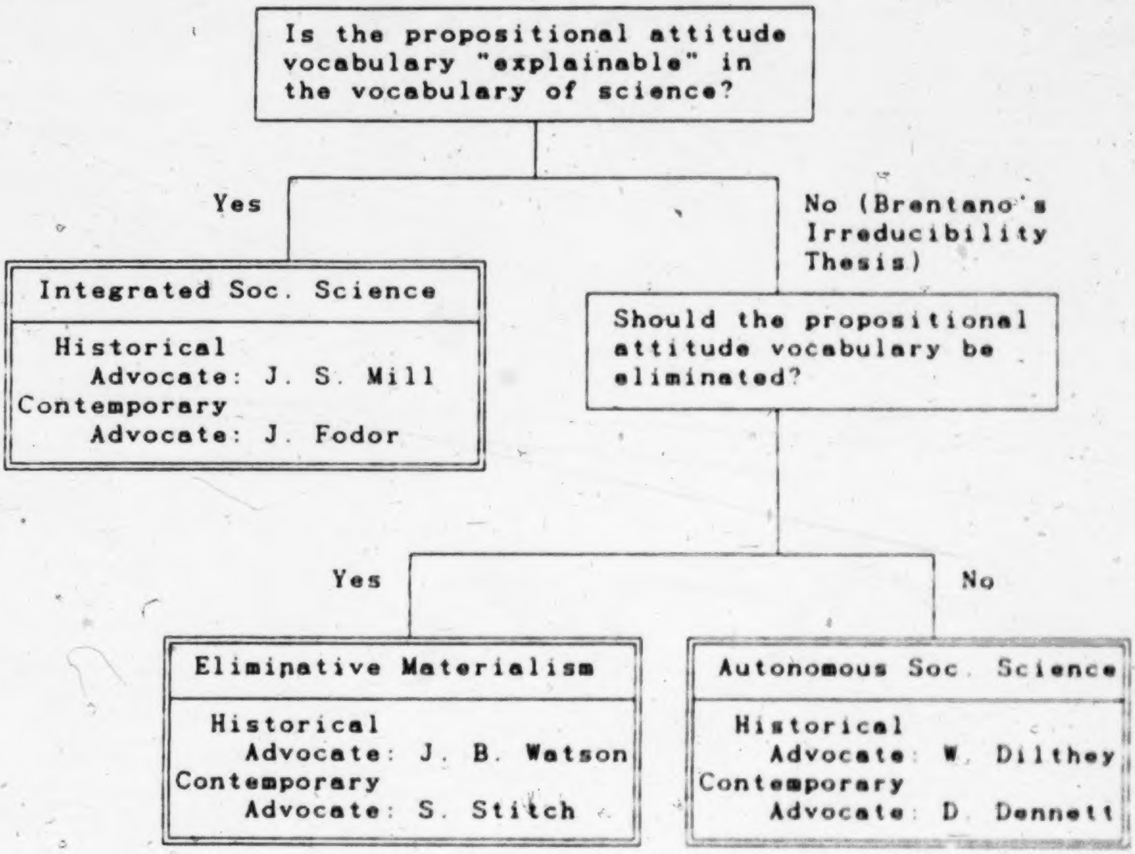
> ...There remains a thesis of Brentano's ... that is directly relevant to our emerging doubts over the propositional attitudes and other intentional idioms. It is roughly that there is no breaking out of the intentional vocabulary by explaining its vocabulary in other terms... One may accept the Brentano thesis either as showing the indispensability of the intentional idioms and the importance of an autonomous science of intuitions, or as showing the baselessness of intentional idioms and the emptiness of a science of intentions.[251]

It has been almost three decades since this passage was written, and over a century since essentially the same issues were debated in Germany. Yet there are still considerable differences of opinion on these issues. In later sections I will discuss recent work in the philosophy of mind that relates to this debate. The following decision tree is a guide to how the issues will be discussed in this chapter.

---

[251] W. V. O. Quine, Word and Object, (The M.I.T. Press, 1960), 220-221.

Figure 3.1.1
Relation of the *Geisteswissenschaften* to the Natural Sciences


Is the propositional attitude vocabulary "explainable" in the vocabulary of science?
Yes
No (Brentano's Irreducibility Thesis)
Integrated Soc. Science
Historical Advocate: J. S. Mill
Contemporary Advocate: J. Fodor
Should the propositional attitude vocabulary be eliminated?
Yes
No
Eliminative Materialism
Historical Advocate: J. B. Watson
Contemporary Advocate: S. Stitch
Autonomous Soc. Science
Historical Advocate: W. Dilthey
Contemporary Advocate: D. Dennett

Now let us relate these general issues about the *Geisteswissenschaften* to the Linguistic Relativity Hypothesis. The Linguistic Relativity Hypothesis has "thought" as its dependent variable. I claim that "thought" should be conceived in terms of propositional attitudes, specifically, beliefs. That is, Whorf's hypothesis should be construed as the claim that one's beliefs are affected by the language one speaks, where beliefs are construed, at first cut, as some sort of relation between a believer and a proposition. Consequently the Linguistic Relativity Hypothesis must be construed as a hypothesis in the Geisteswissenschaften. Now if J. S. Mill and Fodor are correct, then the Linguistic Relativity Hypothesis is also a hypothesis of the natural sciences, since on their view the social sciences are simply a branch of the natural sciences. If Watson and Stitch are correct, the Linguistic Relativity Hypothesis is just nonsense; it has the same scientific status as the hypothesis that witches consort with the devil, i.e., none. Finally, if Dilthey and Dennett are correct, then the Linguistic Relativity Hypothesis is a hypothesis of the autonomous social sciences, and it must therefore be investigated by means of theoretical and methodological considerations that are unique to the social sciences.

I think Dilthey and Dennett are correct, and the main purpose of this chapter is to make that case. This will set the stage for Chapter Four in which I assess the Linguistic Relativity Hypothesis in terms of a method (commonly called 'radical interpretation', and which, to my way of thinking, is a explication of what Dilthey and Weber were getting at in their discussions of Verstehen) that is unique to the social sciences.

However, before engaging in these issues I want to first address a criticism of the way that I have characterized the Geisteswissenschaften. I have claimed that "subjectively meaningful action" is behavior that can be made sense of in terms of the actor's intentional states, and following Frege, I have argued that intentional states should be thought of as propositional attitudes. However, recently a number of writers have suggested that intentionality is not basically propositional in nature; i.e., our thoughts do not (at least not in the central cases) have the function-argument (i.e., subject-predicate) structure that Frege claimed they have. Now if these claims are correct, then my discussion of the options outlined in Figure 3.1.1 falls apart, since this discussion is based almost entirely on recent philosophical work on the topic of propositional attitudes. Consequently, the next section is devoted to an assessment of these advocates of the case that thought is non-propositional.

3.2 NON-PROPOSITIONAL THEORIES OF THOUGHT

Frege believed that thought is propositional in nature. Whorf, on the other hand, believed that thought is essentially a matter of exercising concepts. Whorf had a theory of thought that appears to be non-propositional. Are these two views in conflict?

My view is that if one adopts a propositional theory of thought, then one is licensed to speak of "concepts" as a convenient shorthand. For to say that we think in propositions is to say that our thoughts have the same semantic structure as sentences, and as we saw from our examination of Tarskian truth definitions in Chapter Two, we make sense of the semantics of sentences by constructing axioms that describe the semantics contributions of individual words to the truth-conditions of the sentences in which they appear. One class of words are the predicates; they are assigned extensions (sets of objects) in a Tarskian truth-definition. We can, therefore, identify a person's concepts with the semantics of the predicates that he employs. However, it was noted in the previous chapters that different grammars (i.e., different truth-definitions) with different treatments of predicates can yield the same truth-conditions, so there may be a certain amount of looseness and imprecision when we speak of a person's concepts. Also, as will be pointed out in the next chapter, the notion of a language-independent "conceptual scheme" is not implied by the view that thought is propositional. However, it is reasonable to assume that the propositional view of thought does warrant the attribution of concepts to a person, provided that this attribution is viewed as somewhat underdetermined with respect to the more fundamental attribution of thoughts.

Some authors reject the view that thought is primarily propositional, arguing instead that thought is primarily the exercise of concepts. However, I believe that if one simply stops at that point, without attempting to develop the view that thought is also (derivatively) propositional in nature, then one will have an account of thought that is radically defective. (I will argue for this point shortly.) On the other hand, if one tries to claim that thought is primarily conceptual and derivatively propositional, then one is obligated to come up with an account of how propositional thought is based on conceptual thought. This, in my opinion, gets the cart before the horse. In the philosophy of language there is a famous thesis of Frege's called "the primacy of the sentence". The idea is that the semantic features of sentences are more fundamental that the semantic features of the words that appear in the sentences. I agree with this thesis, and I believe that it carries over into the topic of thought being considered here. If we adopt the view that thought consists in propositions that have concepts as constituents, then we should be viewing thoughts as primary and concepts as secondary, rather than the other way around.[252]

In summary, if we adopt the view that thought is primarily non-propositional, we end up with an unacceptable theory of thought - or so I will argue in this section.

Let us first consider the case for the view that thought is propositional. The case is that if thought were not propositional, then two major intellectual systems would break down. Those systems are folk psychology and epistemology.

---

[252] I will not develop an argument for this claim at this point. In effect the argument is made in Chapter Four where I argue that the method of radical interpretation - i.e., a method that tries to simultaneously determine the propositional attitudes of a community and the truth-conditions of their language - is, for the most part, based on observations of certain propositional attitudes about sentences. Thus, we have a methodological argument for the primacy of the sentence. However, for more extensive discussions of this doctrine see L. Wittgenstein, Philosophical Investigations, (1953, reprint Basil Blackwell, 1976); W. V. O. Quine, Word and Object, (The M.I.T. Press, 1960); M. Dummett, Frege: Philosophy of Language, (Harper and Row, 1973; and J. Wallace, "Only in the Context of a Sentence do Words have any Meaning", in P. A. French, T. E. Uehling, Jr., and H. K. Wettstein, eds. Contemporary Perspectives in the Philosophy of Language, (University of Minnesota Press, 1979)

'Folk psychology' is the currently fashionable term for the "theory" that backs our ordinary explanations of each other's actions. It is, roughly, that a person has beliefs and desires, and that a person will perform those actions that, according to his beliefs, will result in the maximum overall fulfillment of his desires. Thus folk psychology allows us to predict that McX will be in Toronto next week to attend the reading of the will, simply on the basis of what we know about McX's beliefs and desires. On the other hand, an account of thought that focuses totally on concepts[253] is not capable of making these predictions about actions. Thus, folk psychology suggests that propositional theories of thought are capable of forming the basis of a general theory of action. On the other hand, it is not at all clear how non-propositional theories of thought can accommodate the broader requirements of the explanation of action. It is clear that if the implicitly propositional theory of thought that is inherent in our ordinary folk psychology were replaced with a non-propositional account, the predictive value of folk psychology would be lost. Folk psychology would fail to perform any useful function for us and would simply be abandoned.

The other major field that requires a propositional theory of thought is epistemology. In the following passage Quine is talking about empiricism, but his points apply to epistemology in general:

---

[253] Such as the classic work: J. S. Bruner, J. L. Goodnow, and G. A. Austin, A Study of Thinking, (John Wiley and Sons, 1956).

> In the past two centuries there have been five points where empiricism has taken a turn for the better. The first is the shift from ideas to words. The second is the shift of semantic focus from terms to sentences. The third is the shift of semantic focus from sentences to systems of sentences. The fourth is, in Morton White's phrase, methodological monism: the abandonment of the analytic-synthetic dualism. The fifth is naturalism: the abandonment of the goal of a first philosophy prior to natural science.[254]

The first two steps, which Quine dates in the late eighteenth century, constitute the adoption of the propositional model in epistemology. The others were possible only because that model has been adopted (with the fifth being a possible exception). Any survey of epistemological issues[255] will convince one that all the problems are currently formulated in terms of a propositional model of knowledge and belief. Reversion to a non-propositional model of thought would entail rewriting the epistemological progress of the last two centuries.

I submit, then, that the propositional model of thought is warranted because of its integral role in two major intellectual systems: folk psychology and epistemology. The theorist who wants to recommend a non-propositional model of thought therefore has a very major task. That theorist must rebuild those systems in terms of his alternative model. The burden of proof is on the shoulders of the non-propositionalist.[256]

---

[254] W. V. O. Quine, "Five Milestones of Empiricism", in Theories and Things, (Harvard University Press, 1981).

[255] See, for example, J. Dancy, Introduction to Contemporary Epistemology, (Basil Blackwell, 1985).

[256] To be fair, the non-propositionalist has another option. He can deny that folk psychology and epistemology are worth keeping. He will not be lacking in company, for these two intellectual systems have been under attach of late. A recent

Lately there have been a number of writers who have criticized the propositional model of thought. I group these critics into two classes. On the one hand, there are a number of cognitive scientists who argue that the propositional model needs to be supplemented with other forms of mental representation. Such theorists argue, for example, that non-propositional "visual images" must be postulated in addition to propositional beliefs.[257] I will not be concerned with these theorists, as they are doing nothing more than suggesting a few patches to the propositional model. On the other hand, there are theorists who suggest that the propositional model is a big mistake, and that it should be replaced by another model. These theorists seem largely to be linguists who hold the view that linguistics is a branch of psychology.[258] It is these theorists that I wish to briefly address.

---

work that suggests that we should abandon folk psychology is S. Stitch, From Folk Psychology to Cognitive Science, (The M.I.T. Press, 1983). I will discuss, and reject, Stitch's views later in this chapter. An influential work that attacks the enterprise of epistemology is R. Rorty, Philosophy and the Mirror of Nature, (Princeton University Press, 1979). I will not attempt to defend epistemology from Rorty's arguments, but see H. Putnam, "Why Reason Can't Be Naturalized", Synthese, 1982, 52, 1-23.

[257] See O. J. Flanagan, Jr., The Science of the Mind, (The M.I.T. Press, 1984), 187-192; N. A. Stillings, et. al., Cognitive Science, (The M.I.T. Press, 1987), 30-48; and the essays in part two of N. Block, ed. Readings in the Philosophy of Psychology, Volume 2, (Harvard University Press, 1981).

[258] See the annotated bibliography in G. Lakoff, "Cognitive Semantics", in U. Eco, M. Santambrogio, and P. Violi, eds., Meaning and Mental Representations, (Indiana University Press, 1988).

I will examine the views of George Lakoff as representative of the non-propositional approach to thought.[259] Lakoff distinguishes two major competing approaches to cognition. On the one hand, there is the position of objectivist cognition, which adopts a propositional view of thought, and which is characterized by its adherence to the following eight doctrines, which together define what Lakoff calls "objectivist metaphysics":

1. The world consists of entities with fixed properties and relations holding among them at any instant. This structure is mind-free, that is, independent of the understanding of any beings.
2. The entities in the world are divided up naturally into categories called natural kinds. All natural kinds are sets defined by the essential properties shared by their members.
3. All properties are either complex or primitive; complex properties are logical combinations of primitive properties.
4. There are rational relations that hold objectively among the entities and categories in the world. For example, if an entity x is in the category A and if A is in the category B, then x is in B.
5. The doctrine of truth-conditional meaning: Meaning is based on truth and reference.
6. The "correspondence theory" of truth: Truth consists in the correspondence between symbols and states of affairs in the world.
7. The doctrine of objective reference: There is an "objectively correct" way to associate symbols with things in the world.
8. Conceptual categories are designated by sets characterized by necessary and sufficient conditions on the properties of their members.

---

[259] I will concentrate on his recent paper "Cognitive Semantics", op. cit.

9. A complex concept is DEFINED by a collection of necessary and sufficient conditions on less complex (and ultimately, primitive) concepts.[260]

The other approach to cognition, according to Lakoff, is radically different from the objectivist theory. Lakoff supports this alternate theory, which he calls experientialist cognition.

> The theory of experientialist cognition posits:
>
> - Concepts of two sorts that are meaningful because of their roles in bodily experience (especially movement and perception):
>
>   1. Basic level concepts...
>
>   2. Image-schemas (e.g., containers, paths, links, part-whole schemas, force-dynamic schemas, etc.) These have a non-finitary internal structure.
>
> - Imaginative processes for forming abstract cognitive models from these: Schematization, Metaphor, Metonymy, and Categorization.
>
> - Basic cognitive processes such as focusing, scanning, superimposition, figure-ground shifting, vantage-point shifting, etc.
>
> - Mental spaces...
>
> The central claim of experientialist cognition is:
>
> - Meaningful conceptual structures arise from two sources:
>
>   1. from the structured nature of bodily and social experience and
>
>   2. from our innate capacity to imaginatively project from certain will-structured aspects of bodily and interactional experience to abstract conceptual structures.
>
> Rational thought is the application of very general cognitive processes - focusing, scanning,

---

[260] Ibid., these points are distributed between p. 123 and 136.

superimposition, figure-ground reversal, etc. - to such structures.[261]

Rational thought can be propositional in nature, according to Lakoff, but such propositional thought is very much derivative on a much more basic and pervasive non-propositional type of thought.

Lakoff argues that "objectivist metaphysics" is faulty on a number of grounds. Given its unacceptability, argues Lakoff, we must replace it with a philosophical position that is essentially Kantian. Once we recognize the sterility of objectivist metaphysics our new Kantian perspective will lead us straight to the theory of experientialist cognition.

The overall structure of Lakoff's argument is as follows: (1) Objectivist cognition (i.e., the theory that thought is propositional) presupposes objectivist metaphysics; (2) objectivist metaphysics is seriously in error, and should be replaced with a more-or-less Kantian metaphysics; and (3) this new metaphysics will lead us to the alternate theory of experientialist cognition.

I think this argument fails entirely; (1) and (2) are definitely false, and (3) may be as well. However, I do not wish to engage in this debate. I think the Kantian metaphysics that have been adopted by a large number of theorists of late[262] is something of a red herring. Underlying all the Kantianism there is a deeper reason why these theorists are unhappy with propositional accounts of thought. The real reason, I believe, is that there is a growing suspicion that the propositional account of thought is not the proper foundation on which to build a scientific theory of thought. So let us put the metaphysical issues to one side and concentrate on the more interesting complaint, viz., that a scientific theory of thought requires some other view of thought than that it is propositionally based.

---

[261] Ibid., 121.

[262] Besides Jackendoff and Lakoff, see G. Fauconnier, Mental Spaces: Aspects of Meaning Construction in Natural Language, (The M.I.T. Press, 1984).

I think that this concern can best be appreciated when we stop to review the recent history of artificial intelligence research. Much (but not all) of the work in artificial intelligence can be characterized as an attempt to replicate certain cognitive functions, and it is commonly held that in order to do this fragments of "knowledge" relevant to the task at hand must be represented in the computer. At first the propositional model was the most commonly used.[263] However, as artificial intelligence researchers moved from relatively simple problems, such as theorem proving, to more difficult tasks, such as natural language processing, they became dissatisfied with propositional models of knowledge representation. The newer models, with names like "scripts", "schemas", and "frames" were more complex data-structures intended to represent a "stereotype" of a situation. Marvin Minsky sums up the approach as follows:

---

[263] For example, the programs that were based on the "resolution principle" (a rule of inference for languages with the structure of first order predicate logic) developed by J. A. Robinson. Such programs are described in J. R. Slagle, Artificial Intelligence: The Heuristic Programming Approach, (McGraw-Hill, 1972), chapter 5; and P. C. Jackson, Introduction to Artificial Intelligence, (Petrocelli/Charter, 1974), chapter 6.

> Here is the essence of the frame theory: When one encounters a new situation (or makes a substantial change in one's view of a problem), one selects from memory a structure called a frame. This is a remembered framework to be adapted to fit reality by changing the details as necessary.
>
> A frame is a data-structure for representing a stereotyped situation like being in a certain kind of living room or going to a child's birthday party. Attached to each frame are several kinds of information. Some of this information is about how to use the frame. Some is about what one can expect to happen next. Some is about what to do if these expectations are not confirmed.
>
> We can think of a frame as a network of nodes and relations. The 'top levels' of a frame are fixed, and represent things that are always true about the supposed situation. The lower levels have many terminals - 'slots' that must be filled by specific instances or data.[264]

The real motivation for incorporating data structures like these into artificial intelligence research, I believe, is that they "are used to make certain kinds of calculations economical".[265] Computer programmers solve their problems by means of setting up two elements, a set of data structures used to store information relevant to the problem, and an algorithm (i.e., a program) that operates on those data structures (and the various input/output devices of the computer system). Over the years, computer scientists have developed a large number of data structures (e.g., elementary structures such as stacks, queues, linked lists, trees, etc., and more complex structures built from these elementary structures) that are optimized for particular problems.[266] Choosing the right data structure for a problem can greatly reduce the complexity and length of execution of the algorithm needed to solve the problem. Given this standard tactic of computer programmers, is it any wonder that artificial intelligence researchers have developed data structures (e.g., frames) that are optimized for the particular computational problems that they have set for themselves? However, there is a tendency for the proponents of frames, scripts, etc., to turn necessity into a virtue, and to view their data-structures as realistic models of human mental representation. Thus Minsky writes:

---

[264] M. Minsky, "Frame-system Theory", in P. N. Johnson-Laird, and P. C. Wason, eds., Thinking: Readings in Cognitive Science, (Cambridge University Press, 1977), 355.

[265] Ibid., 355.

> It seems to me that the ingredients of most theories both in Artificial Intelligence and in Psychology have been on the whole too minute, local, and unstructured to account - either practically or phenomenologically - for the effectiveness of common-sense thought. The "chunks" of reasoning, language, memory, and perception ought to be larger and more structured; their factual and procedural contents must be more intimately connected in order to explain the apparent power and speed of mental activities.
>
> Similar feelings seem to be emerging in several centers working on theories of intelligence... I see all these as moving away from the traditional attempts both by behavioristic psychologists and by logic-oriented students of Artificial Intelligence [i.e., those who use a propositional model of thought] in trying to represent knowledge as collections of separate, simple fragments.[267]

---

[266] For a typical survey, see E. Horowitz, and S. Sahni, Fundamentals of Data Structures in Pascal, (Computer Science Press, 1984).

[267] M. Minsky, "A Framework for Representing Knowledge", in J. Haugeland, ed., Mind Design, (The M.I.T. Press, 1981), 95.

A key shift takes place in this passage when Minsky argues that the frame theory is preferable from a phenomenological perspective as well as from a practical perspective. Frame theory is a better approach, according to Minsky, because it describes how people actually think.

From this it is a short step to rejecting the propositional model as a model of how people think. Thus, in a recent work, Minsky argues that even our knowledge of language - which is propositional if anything is - is best accounted for in terms of frame theory.[268] In other words, frame theory is not seen as an adjunct to the propositional model; it is seen as the fundamental account of mental representation, upon which the propositional model, insofar as it is required, can be derived.

Recent work in "experientialist cognition" - e.g., Mark Johnson's claim that all cognition is ultimately based on kinaesthetic body-schemas such as "the container schema", "the part-whole schema", etc.[269] - can be seen as a further development and elaboration of Minsky's idea.

We can summarize these observations as follows. Of late, some people have expressed dissatisfaction with the propositional model of thought. Sometimes they claim that their dissatisfaction with that model is based on their "realization" that a Kantian epistemology is the only way to go. However, another source of dissatisfaction is the view that a propositional model of thought will not yield a valid scientific account of cognition.

---

[268] M. Minsky, The Society of Mind, (Simon and Schuster, 1985), 261-272.

[269] M. Johnson, The Body in the Mind: The Bodily Basis of Reason and Imagination, (University of Chicago Press, 1987).

As mentioned above, I think the Kantian argument is best put to one side. But what are we to make of the other argument? My response is very simple. I accept that the propositional model of thought is not a sound base for a scientific theory of cognition. But I reject the idea that we should therefore abandon the propositional model of thought for an alternate theory, say frame theory, or kinaesthetic body-schema theory. Instead, I think we should give up the quest for a scientific theory of cognition. But I must be more precise. We should give up trying to build a scientific theory of cognition that deals with "content", i.e., with the intentionality (in Brentano's sense) of our mental life. On the other hand, a scientific theory of cognition that deals only with "information" (as opposed to "content") is perfectly acceptable. Such a theory, however, would fail to be of interest to Brentano, and for that matter, to Whorf.

The argument against the quest for a scientific theory of content will be presented shortly. For now, assume that there is something deeply wrong with this quest. Then the arguments of Minsky and Johnson (as I have interpreted him) against the propositional model no longer have any force. Why should I give up the propositional model of thought on the grounds that it is inappropriate for a scientific theory of content, if the very idea of a scientific theory of content is itself incoherent?

The comeback, of course, is as follows: If the propositional model of thought cannot be made scientific, what grounds do we have for holding on to that model? This is exactly the issue between the eliminative materialists and the advocates of an autonomous social science (see figure 3.1.1). In later sections of this chapter I will attempt to argue for the autonomous social science position. But notice that if that position is correct, then Minsky's argument for a non-propositional model of thought will have been thoroughly undermined.

So let us assume, then, that the imposing edifices of folk psychology and epistemology provide sufficient warrant for the propositional model of thought. I will now go on to argue that (1) this model cannot be reduced to "hard" science, and (2) nonetheless, we should retain that model in an "autonomous" social science. If I am correct in these arguments I will have also undermined whatever motivation one might have for developing a non-propositional model of thought.

3.3 REDUCIBILITY, SUPERVENIENCE, AND PROJECTIBILITY

Weber plus Brentano plus Frege equals the idea of a social science grounded in the notion of rational action, where rational action is understood as comprising those acts (bodily movements) that people perform as a "rational function" of their beliefs and desires, where beliefs and desires are understood as propositional attitudes. When I say "rational function" I mean that the agent chooses acts (bodily movements) that are, according to his beliefs, the best way of achieving his desires.[270] The question now to be considered is the first question from figure 3.1.1: Is this conception of the social sciences compatible with natural science? The main purpose of this section is to address some of the preliminary issues surrounding this question.

Philosophical Behaviorism

Propositional attitude psychology would fit quite nicely into a scientific world view if it were possible to *reduce* the unique terms of propositional attitude psychology to the terms of physics, or barring that, to the terminology of some descriptive level that can itself be reduced to physics.

Reduction will be (crudely) characterized as follows. Theory A reduces to theory B if, for every predicate P of A, there is a true "bridge law" of the form:

For all x, Px iff Qx

where Q is a primitive predicate, or a fairly simple complex predicate (i.e., a logical compound of primitive predicates) of Q.

---

[270] This is very vague, of course. It can be sharpened by selecting a more precisely defined function from decision theory, such as maximization of expected utility. However, since there are many such functions in decision theory, it is an open question as to which of these models is most appropriate to ordinary human action. For discussions, see C. G. Hempel, *Aspects of Scientific Explanation*, (The Free Press, 1965), 463-469.

It is now quite apparent that it is not possible to reduce propositional attitude psychology to a more basic discipline. A number of Anglo-American philosophers did attempt such a reduction in the years after World War II. This attempt has come to be known as philosophical behaviorism, and the main tactic was to view propositional attitudes as dispositions to behavior. There are a number of significant criticisms of this attempted reductions. Two of the most-telling criticisms are given in the following passage by Paul Churchland.

> According to the [philosophical] behaviorist ... to say that Anne wants a Caribbean holiday is to say that (1) if asked whether that is what she wants, she would answer yes, and (2) if given new holiday brochures for Jamaica and Japan, she would peruse the ones for Jamaica first, and (3) if given a ticket on this Friday's flight to Jamaica, she would go, and so on and so on...
>
> The list of conditionals necessary for an adequate analysis of "wants a Caribbean holiday" ... seemed not just to be long, but to be indefinitely or even infinitely long, with no finite way of specifying the elements to be included. And no term can be well-defined whose definiens is open-ended and unspecific in this way. Further, each conditional of the long analysis was suspect on its own. Supposing that Anne does want a Caribbean holiday, conditional (1) above will be true only if she isn't secretive about her holiday fantasies; conditional (2) will be true only if she isn't bored with the Jamaica brochures; conditional (3) will be true only if she doesn't believe the Friday flight will be hijacked, and so forth. But to repair each conditional be adding in the relevant qualification would be to reintroduce a series of mental elements into the business end of the definition, and we would no longer be defining the mental solely in terms of publicly observable circumstances and behavior.[271]

---

[271] P. M. Churchland, Matter and Consciousness, (The M.I.T. Press, 1984), 23-24.

The Neo-Wittgensteinian Argument

From this point on, I will assume that philosophical behaviorism is false, and all the theorists I will discuss share that assumption. Our question now is: Can the Weberian concept of social science be made compatible with natural science in some manner other than reduction? In the late fifties and early sixties a number of philosophers under the influence of the later work of Wittgenstein argued that the answer to this question is in the negative.[272] This neo-Wittgensteinian response was based on a supposed distinction between reasons and causes. The argument goes as follows: Normally when scientists develop theories that deal with dynamical phenomena (i.e., events in time), they look for causal relations between events. Newtonian mechanics is a paradigm example of this kind of causally based science. Now at first glance, explaining a bit of bodily motion by giving the agent's reason for moving in that way *appears* to be a causal explanation of a dynamical phenomenon: e.g., he snatched the purse *because* he wanted the money. However, appearances are deceiving, according to the neo-Wittgensteinians. Genuine causal relations in science are contingent empirical relations. A scientific hypothesis that one class of events (say, the presence of HIV) causes another (the development of AIDS) must be supported with a wealth of evidence, including evidence that rules out other possible causative agents. Only then can we speak of causality. But this is not how reasons seem to be related to actions. The agent knows the reasons for his action immediately and directly; not as the result of a statistical analysis of his previous mental and physical activity. Secondly, reasons and actions are not conceptually independent the way that cause and event are supposed to be in scientific explanations. For if a person has never taken any action to acquire money, is it possible to say of that person that he wants money? The answer (according to the neo-Wittgensteinians) is no: wanting money is a disposition to acquire money where feasible. So given this differences between reasons and paradigm cases of causality in science, we must conclude that reasons are not mental events that precede and cause actions. What are reasons then? The neo-Wittgensteinian answer is that to give a reason for an action is to redescribe that action in a special vocabulary, the vocabulary of human goals and objectives. Peter Winch went on to point out that this neo-Wittgensteinian doctrine has obvious implications for the Weberian social science we are considering; specifically, the distinction between reasons and causes implies that

---

[272] Two examples are P. Winch, *The Idea of a Social Science*, (Routledge and Kegan Paul, 1958); and A. I. Melden, *Free Action*, (Routledge and Kegan Paul, 1961). For a more extensive bibliography of this position, see K. Donnellan, "Reasons and Causes", in P. Edwards, ed. *The Encyclopedia of Philosophy*, (MacMillan, 1967), Volume 7, 88; and A. I. Goldman, *A Theory of Human Action*, (Princeton University Press, 1970), 77

> . . . the conceptions according to which we normally think of social events are logically incompatible with the concepts belonging to a scientific explanation.[273]

So, in the wake of Wittgenstein's influence it appeared that the social studies (or more precisely, a Weberian style of social theory) could not be integrated into the natural sciences. However, this neo-Wittgensteinian argument was soon subjected to a major critique in a classic paper by Donald Davidson. Writing in 1963, Davidson stated that the purpose of his paper was

---

[273] P. Winch, op. cit., 95.

> ... to defend the ancient - and commonsense - position that rationalization is a species of causal explanation. This defence no doubt requires some redeployment, but it does not seem necessary to abandon the position, as has been urged by some recent writers.[274]

I will not discuss all the subtleties of Davidson's paper; rather, I will restrict myself to his comments on the two neo-Wittgensteinian objections to identifying reasons with causes that were raised above. One point raised by the neo-Wittgensteinians was that reasons are conceptually related to actions, and therefore could not be the causes of actions, because causes and effects have to be conceptually distinct. Davidson points out that this is a mistake; if the impact of billiard ball A causes the motion of billiard ball B, then the impact of billiard ball A can be redescribed as the cause of the motion of billiard ball B. Thus we can say that the cause of the motion of billiard ball B caused the motion of billiard ball B, and thereby identify a genuine contingent causal relation, in spite of the "conceptual relation" between the way we have described the cause and the effect.

The other neo-Wittgensteinian point was the claim that causal relations are based on the observation of a large number of cases, whereas a typical first person rationalization does not depend on extrapolation of from a number of observations. Now

---

[274] D. Davidson, "Actions, Reasons, and Causes", Journal of Philosophy, 1963, 60, 685-700. Reprinted in D. Davidson, Essays on Actions and Events, (Oxford University Press, 1980), 3.

Davidson does not disagree that a causal relation must be grounded in a large number of cases. Another way of putting this is that a causal relation must be supported by a causal law, where a causal law is a generalization of the large number of observations of concomitant events. Let us call the principle that a singular causal explanation (e.g., the statement that the impact of billiard ball A caused the motion of billiard ball B) must be supported by a causal law (e.g., Newton's laws of motion) "Hume's principle". The neo-Wittgensteinian point, then, is that if reasons are the causes of actions then, by Hume's principle, giving a reason must be an instance of what is known as deductive-nomological explanation; i.e., explaining an event as an instance of a general law. But when we give reasons we do not seem to be invoking causal laws at all. Therefore, conclude the neo-Wittgensteinians, to give a reason for an action is not to give a cause.

Davidson shows that it is possible to (1) adhere to Hume's principle, (2) agree with the neo-Wittgensteinians that when we give reasons we typically do not overtly invoke laws, and (3) still maintain that reasons are causes. In explaining Davidson's argument I will invoke the distinction between supervenience and reducibility, a distinction that will be important for later arguments in this work. I will therefore interrupt the main thread of the argument to discuss these concepts.

Reducibility and Supervenience

The terms 'supervenience' and 'reducibility' have come to be used to distinguish two senses in which two different explanations of the same events can be related to one another.[275] The distinction between supervenience and reducibility can be demonstrated by examples. Consider the relation between Mendelian genetic explanations and biochemical explanations. Note that (1) every Mendelian event is strictly identical with a biochemical event, and (2) every explanatory principle of Mendelian genetics is just a "summary" of certain biochemical principles. Now consider the relation between a historical explanation (the kind found in history books) and a physical explanation. Note that (3) every historical event is identical with (a) physical event(s), but (4) it is not the case that the principles of historical explanation are "summaries" of the principles of physics, indeed, it is not even clear that there are general lawlike principles of historical explanation. Now, in light of (3) I claim that historical explanations are supervenient on physical explanations, but in light of (4) I claim that historical explanations are not reducible to physical explanations. In contrast, Mendelian explanations are both supervenient on and reducible to biochemical explanations.

More exactly, a set of properties M supervenes on a set of properties N with respect to a domain D only if it is necessarily the case that any objects of D which share all their N#

---

[275] As in J. Kim, "Supervenience and Nomological Incommensurables", American Philosophical Quarterly, 1978, 15, 149-156.

properties also share all their M# properties, where M# and N# are the closures of M and N, respectively, under all operations (e.g., conjunction, disjunction) whereby properties are formed from properties.[276] An explanation A supervenes on an explanation B only if every predicate used in A expresses a property which is a member of a set of properties X which supervenes on a set of properties Y such that every predicate used in B expresses a property which is a member of Y. The other term is 'reducibility', which I am using in Nagel's sense: i.e., a law of science S_M reduces to the laws of S_N only if there are lawlike "bridge principles" such that the S_M law is formally deducible from the S_N laws and the bridge principles.[277] Philosophical behaviorism was an attempt to establish such bridge principles between propositional attitude psychology and physics.

The contrast between supervenience and reducibility has long been hidden, I believe, by the once dominant model of scientific explanation, the deductive-nomological (DN) model.[278] Explanation, according to this model, is a relation between sentences. If E is the sentence reported the state of affairs to be explained then, according to the model, an explanation takes the form

---

[276] Ibid., 152.

[277] E. Nagel, The Structure of Science, (Harcourt, Brace and World, 1961), chapter 11.

[278] Cf. C. G. Hempel, Aspects of Scientific Explanation, (The Free Press, 1965), 335-37.

$$\begin{array}{l} C_1, \ldots, C_n \\ L_1, \ldots, L_m \\ \hline E \end{array}$$

where $C_1, \ldots, C_n$ are sentences describing specific individual conditions, $L_1, \ldots, L_m$ are scientific laws, and the horizontal line indicates that the E is logically deducible from the initial condition statements and the laws. On this account, 'C explains E' is an elliptical version of a deduction of the above form, i.e., elliptical in the sense that the relevant laws have not been mentioned.

Let us now examine why supervenience and reducibility are conflated in this model. Suppose that 'C_M explains E_M' is an elliptical explanation in the sense just explicated. The subscripts indicate that every predicate employed in that sentence expresses a property for the set of properties named by that subscript. Now suppose that M is supervenient on N, another set of properties. Suppose further that there is an elliptical explanation 'C_N explains E_N' such that the predicates used in C_M and C_N apply to the same initial conditions; likewise the predicates used in E_M and E_N apply to the same state of affairs. The situation can be (loosely) represented as follows:

Figure 3.3.1
Two Sciences Explaining the Same Causal Relation

$$\begin{tikzcd}
C_M \arrow[r] & \text{State of Affairs \#1} & C_N \arrow[l] \\
\underline{L_M} & & \underline{L_N} \\
E_M \arrow[r] & \text{State of Affairs \#2} & E_N \arrow[l]
\end{tikzcd}$$

Now, what is the relation between the extensions of the predicates employed in the two sets of laws, L_M and L_N? Either the predicates of L_M and L_N carve up the events they describe in fundamentally the same way or they do not. If they do not, then we seem to be faced with a miracle: somehow two fundamentally incongruent ways of categorizing the events of the world yield perfectly harmonious predictions. It would seem, if we temporarily suspend our Humean principles and indulge in a bit of reification, that there are two sets of independent nomological forces which - miraculously - generate state #2 from state #1. But the postulation of two harmonious sets of forces (or rather, the adoption of two incongruent sets of predicates make it seem if there are two sets of forces) is in violation of a venerable epistemic norm: Occam's Razor. Therefore there is only one set of forces, which L_M and L_N encode differently. This means that there must be a systematic relation between M-terms and N-terms. And given a specification of this relation (i.e., the bridge laws) we will be able to deduce L_M from L_N (assuming S_N to be the more basic science). And this is just what reducibility amounts to. In other words, given the assumptions of the DN model of explanation and adherence to Occam's Razor, supervenience implies reducibility. Since reducibility implies supervenience as a consequence of the way they are defined, we see that the DN model ends up conflating the two notions.

Given that I endorse the distinction between supervenience and reducibility I am under some obligation to say what is wrong with the DN model. The appeal of the DN model is that it does justice to what I have called "Hume's principle"; i.e., the principle that a causal relation holds between two events only if there is a general causal law relating events of those types. However, this principle can be interpreted in two ways. There is the strong reading, on which 'C_M caused E_M' presupposes that there is a set of laws L_M such that C_M and L_M imply E_M. (Recall that by our convention regarding subscripts L_M is a set of laws which are expressed in the same predicates that were employed in C_M and L_M.) Clearly, the strong reading is exactly what is claimed in the DN model, for the DN model implies that 'C_M caused E_M' is elliptical for 'C_M and L_M imply E_M'.

However, there is also a weak reading of Hume's principle. On the weak reading 'C_M caused E_M' presupposes that there are sentences C_N and E_N extensionally equivalent to C_M and E_M respectively,[279] and that there is a law L_N (which is expressed in the same predicates as C_N and E_N) such that C_N and L_N imply E_N. But it is not necessarily the case that there is a set of laws L_M that "backs" the original M-explanation. In other words, the weak reading of Hume's principle is that a causal explanation must be backed by a causal law, but it is not necessary that the causal law be formulated in the same predicates as the causal explanation. The two readings are represented in the following two figures.

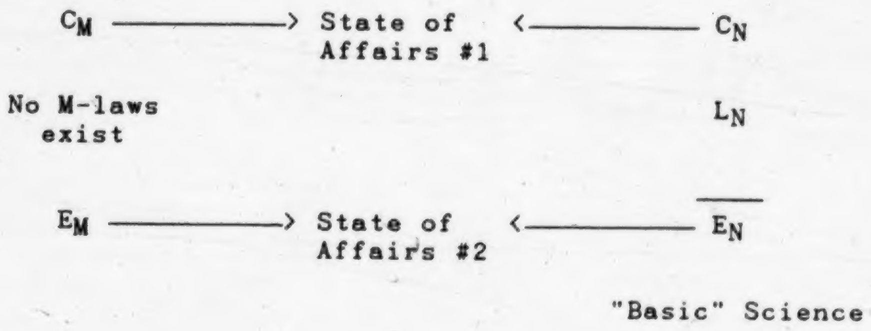
---

[279] Extensionally equivalent in the sense that the referring phrases of the pairs of sentences pick out the same objects and sets of objects in the world. For the sake of brevity I am being somewhat imprecise in my formulations.

Figure 3.3.2
Two Sciences Explaining the Same Causal Relation
According to the "Strong" Reading of Hume's Principle

$$\begin{tikzcd}
C_M \arrow[r] & \text{State of Affairs \#1} & C_N \arrow[l] \\
L_M \arrow[rr, dashed, "\text{Nagelian ``reduction'' via bridge laws}"'] & & L_N \\
\overline{E_M} \arrow[r] & \text{State of Affairs \#2} & \overline{E_N} \arrow[l] \\
 & & \text{``Basic'' Science}
\end{tikzcd}$$

Figure 3.3.3
Two Sciences Explaining the Same Causal Relation According to the "Weak" Reading of Hume's Principle



Although the strong reading is valid in cases where one science has been successfully reduced to another more basic science, it is quite apparent that the weak reading is the correct one for many causal explanations. Davidson makes the point in the following passage:

> The most primitive explanation of an event gives its cause; more elaborate explanations may tell more of the story, or defend the singular causal claim by producing a relevant law or by giving reasons for believing such exists. But it is an error to think no explanation has been given until a law has been produced. Linked with these errors is the idea that singular causal statements necessarily indicate, by the concepts they employ, the concepts that will occur in the entailed law. Suppose a hurricane, which is reported on page 5 of Tuesday's *Times*, causes a catastrophe, which is reported on page 13 of Wednesday's *Tribune*. The event reported on page 5 of Tuesday's *Times* caused the event reported on page 13 of Wednesday's *Tribune*. Should we look for laws relating events of these *kinds*? It is only slightly less ridiculous to look for a law relating hurricanes and catastrophes. The laws needed to predict the catastrophe with precision would, of course, have no use for concepts like hurricane and catastrophe. The trouble with predicting the weather is that the

> descriptions under which events interest us - 'a cool, cloudy day with rain in the afternoon' - have only remote connections with the concepts employed by the more precise known laws.[280]

To conclude this discussion of the distinction between supervenience and reducibility, note that where one field of study reduces to physics (where presumably the real story of causation is told), then that field of study will have real causal laws to back its singular causal statements. However, a field of study may fail to reduce to physics and instead supervene on physics. In this case it will not have its "own" causal laws. However, it is still possible to formulate singular causal statements within such a discipline, however, they will be "backed" by causal laws formulated in an entirely different vocabulary.

Davidson's Refutation of The Neo-Wittgensteinians

Now to return to the topic at hand. The neo-Wittgensteinian argument was that reasons could not be causes because causal explanations require backing laws, and these appear to be absent in typical rationalizations. However Davidson has shown how, on the weak reading of Hume's principle, the requisite backing laws can come from a different science. He concludes that psychological explanations and physical explanations are related in exactly this way, and that, after all, reasons are causes.

However, in spite of the defence of the causal nature of rationalizations, Davidson's paper does not lead to the view that the social sciences are reducible to the natural sciences. Rather, his argument is that there are no laws of rational action, and that the whole vocabulary of actions, beliefs, desires, etc., is merely supervenient on the vocabulary of physics.[281]

---

[280] D. Davidson, op. cit., 17.

However, Davidson's paper raises issues that not only undermine the scientific status of explanations by reasons, they also work to undermine the very coherence of the intentional vocabulary in which those explanations are couched. To see why this is so we must first become acquainted with a philosophical issue formulated by Nelson Goodman: his "new riddle of induction".

Goodman on Projectibility

Goodman was interested in an epistemological problem: induction. The particular problem he is interested in is brought out by the following puzzle:

> Suppose that all emeralds examined before a certain time t are green. At time t, then, our observations support the hypothesis that all emeralds are green; and this is in accord with our definition of confirmation. Our evidence statements assert that emerald a is green, that emerald b is green, and so on; and each confirms the general hypothesis that all emeralds are green. So far, so good.
>
> Now let me introduce another predicate less familiar than "green". It is the predicate "grue" and it applies to all things examined before t just in case they are green but to other things just in case they are blue. Then at time t we have, for each evidence

---

[281] Davidson calls his position "anomalous monism". The term 'anomalous' highlights Davidson's view that there are no strict psychological laws. The term 'monism' indicates that Davidson is a materialist, and rejects Cartesian dualism.

> statement asserting that a given emerald is green, a parallel evidence statement asserting that that emerald is grue. And the statements that emerald a is grue, emerald b is grue, and so on, will each confirm the general hypothesis that all emeralds are grue.[282]

But there is obviously something wrong with the induction to the claim that all emeralds are grue. Clearly, Hume's account of induction, which is simply that we should observe past regularities, is inadequate because it permits the offensive induction. As Goodman puts it, "to say that valid predictions are those based on past regularities, without being able to say which regularities, is thus quite pointless."[283] Goodman's new riddle of induction, then, is to specify criteria that would rule out bogus predicates, such as 'grue', as bases for induction.

Goodman coins the term 'projectible' to apply to predicates which, unlike 'grue', support genuine lawlike generalizations. The new riddle of induction can therefore be reformulated as the quest for a theory of projectibility, for what we are after is an account of which predicates allow us to formulate hypotheses that have a chance at being laws of nature. Goodman does not claim to have solved the problem of projectibility, but he has provided a sketch of a theory of projectibility that he thinks will work. It is summed up in the following passage by Israel Scheffler:

> Goodman's suggestion is to make use of pragmatic or historical information that may fairly be assumed available at the time of induction, and to define projectibility in terms of such extra-syntactic, extra-semantic information... Goodman ... suggests that, in order to specify the generalizations chosen on given

---

[282] N. Goodman, Fact, Fiction, and Forecast, 4th ed., (Harvard University Press, 1983), 73-74.

[283] Ibid., 82.

> evidence, we need not restrict ourselves exclusively to the non-pragmatic features of the statements before us. Rather, he proposes that we use also the historical record of past predictions, or (more generally) projections, and in particular, the biographies of the predicates previously used projectively. Our hypotheses, he suggests, are chosen not by virtue of the way they encompass the evidence, but also by virtue of the way the language in which they are couched conforms with past linguistic practise.
>
> His basic term is 'entrenchment', applicable to predicates in the degree to which they (or their extensional equivalents, that is, words picking out the same class of elements, e.g., 'triangle' and 'trilateral') have actually been previously employed in projection: in formulating particular inductive judgments.[284]

So Goodman's theory is that a predicate is suitable for forming inductive generalizations, i.e., it is projectible, if its extension somehow "conforms" with previously used predicates that have been useful in forming inductive generalizations. Thus a good induction is not only a matter of shrewd observation of nature, it is a matter of playing the linguistic game of one's speech community. Goodman summarizes:

> If I am at all correct, then, the roots of inductive validity are to be found in our use of language. A valid prediction is, admittedly, one that is in agreement with past regularities in what has been observed; but the difficulty has always been to say what constitutes such agreement. The suggestion I have been developing here is that such agreement with regularities in what has been observed is a function of our linguistic practices. The line between valid and invalid prediction (or inductions or projections) is drawn upon the basis of how the world is and has been described and anticipated in words.[285]

---

[284] I. Scheffler, The Anatomy of Inquiry, (Bobbs-Merrill, 1963), 311.

[285] N. Goodman, op. cit., 120-121.

If projectible predicates must conform to previous linguistic usage, does this mean that Goodman is disallowing conceptual novelty in science? No, new predicates are allowed, within the following guidelines: Let us define the relation parent of, which holds between pairs of predicates. P is a parent of Q if among the classes that P applies to is the extension of Q.[286] Thus, 'is an animal' is a parent of 'is a tiger'. Goodman then claims that new predicates inherit the entrenchment of their parents. Thus, conceptual progress is allowed so long as entrenchment of the new concepts is conferred by pedigree. But since physics has emerged as our most general and respectable science, this means that new predicates will be projectible to the extent that they are reducible (for Goodman's "parentage" is basically a reductive relation) to physics. Concepts that do not reduce to physics are bad concepts, on a par with 'grue'.

Projectibility and Propositional Attitudes

Now we can see the problem facing Davidson. In arguing that the propositional attitudes do not reduce to the concepts of physics, it appears that Davidson has committed himself to the position that the "mental" concepts we use to characterize propositional attitudes are not projectible - or at least that is what Goodman's doctrine implies. And if these concepts are not projectible - if they are on a par with 'grue' - why do we want to use them at all?

---

[286] Ibid., 106.

Davidson's response is that mental concepts are, in fact, projectible, with a slight caveat. The caveat is that we must not assume that all our concepts must form one mutually entrenched mass. Instead, Davidson suggests that our physical concepts form one island of entrenchment, whereas our mental concepts form another. The two sets of concepts are not entrenched in the same way, and consequently, when psychological reductionists try to formulate bridge laws to reduce the mental to the physical, their reductive statements have all the problems of 'All emeralds are grue'. Davidson develops this argument in the following passage:

> 'All emeralds are green' is lawlike in that its instances confirm it, but 'All emeralds are grue' is not... Nelson Goodman has suggested that this shows that some predicates, 'grue' for example, are unsuited to laws (and thus a criterion of suitable predicates could lead to a criterion of the lawlike). But it seems to me that the anomalous character of 'All emeralds are grue' shows only that the predicates 'is an emerald' and 'is grue' are not suited to one another. Grueness is not an inductive property of emeralds. Grueness is however an inductive property of entities of other sorts, for instance of emerires. (Something is an emerire if is examined before t and is an emerald, and otherwise is a sapphire.) Not only is 'All emerires are grue' entailed by the conjunction of lawlike statements 'All emeralds are green' and 'All sapphires are blue', but there is no reason, as far as I can see, to reject the deliverance of intuition, that it is itself lawlike. Nomological statements bring together predicates that we know a priori are made for each other - know, that is, independently of knowing whether the evidence supports a connection between them. 'Blue', 'red', and 'green' are made for emeralds, sapphires, and roses; 'grue', 'bleen', and 'gred' are made for sapphalds, emerires, and emeroses.
>
> The direction in which the discussion seems to be headed is this: mental and physical predicates are not made for one another. In point of lawlikeness, psychophysical statements [this is Davidson's term for sentences expressing supposed reductions of the mental

> to the physical] are more like 'All emeralds are grue' than like 'All emeralds are green'.[287]

The lesson from Goodman is that a predicate has to belong to a family of similarly entrenched predicates is it is to be useful as a predicate at all. The lesson from Davidson is that we need not assume that there is just one big happy family; there are two important families of concepts, the mental and the physical. However, in light of our discussion of Goodman's views on induction, we can now appreciate that Davidson's Brentanian position poses a delicate problem. Goodman's theory seems to entail that it is success in participating in lawlike statements that gives a predicate its usefulness. In claiming that the mental predicates are a family we are saying that they are mutually entrenched, i.e., that they are projectible in terms of each other, and that seems to mean, according to Goodman's account of induction, that they can be used to formulate hypotheses that will be lawlike if confirmed by their instances. But if there are hypotheses couched in mental predicates that are lawlike, and if we assume a materialistic position with respect to mental events, then it would seem that the only way that we can reconcile the nomological character of the mental with the nomological character of the physical is to assume that the former is reductively related to the later. In other words, Goodman's account of what constitutes a valid family of predicates seems to drive us to the position diagrammed in figure

---

[287] D. Davidson, "Mental Events", in L. Foster and J. W. Swanson, eds., Experience and Theory, (University of Massachusetts Press, 1970). Reprinted in D. Davidson, Essays on Actions and Events, (Oxford University Press, 1980), 218.

3.3.2. But, as we have argued previously, reduction of the mental to the physical is not possible.

What we require, if we are to avoid falling back to the reductionist position, is an account, different from Goodman's, of how a family of predicates can be viable. We need an account of how a family of predicates, and in particular, the family of terms that we use to make propositional attitude ascriptions, can be useful to us in spite of the fact that they do not lead to lawlike hypotheses that comport with the laws of physics.

In the next two sections I will consider two approaches to this problem. In section 3.4 I will consider the philosophy of psychology known as "functionalism". Functionalism is not reductionist, however, it does provide a way of viewing propositional attitude psychology as at least in some sense integrated with the natural sciences. However, I will reject functionalism. In section 3.5 I will consider, and support, the "instrumentalist" position adopted by Daniel Dennett. Dennett's position allows us to maintain the view that we have been advancing; viz., that the social sciences are supervenient on but not reducible to the physical sciences. However, it does not face the formidable problems of functionalism. Instrumentalism, as will shall see, also leads to the position that the social sciences are very much autonomous from the physical sciences.

3.4 FODOR'S FUNCTIONALISM

Let us briefly review the problem developed in the last section.

In the last section we considered the problem of projectibility. Goodman's explanation of this phenomenon is that a predicate is projectible if it is well entrenched with respect to other predicates that have been successfully used, in the past, to form lawlike statements. Now obviously, the predicates of physics are the predicates that have been most successful in this way. Now if some new predicate, say a psychological predicate, is well entrenched with respect to physical predicates, it means that the extension of the psychological predicate is something that can be expressed by means of a fairly compact open sentence employing the predicates of physics. But a sentence stating how the extensions of psychological predicates are related to physical predicates would be a "bridge law" typical of an intertheoretic reduction. Therefore Goodman's position entails the following: A predicate is well-behaved (i.e., projectible) only if it is entrenched with respect to the predicates of physics. But being well-entrenched means that bridge laws can be formulated. Thus a set of predicates, say the psychological predicates, will be well-behaved only if the regularities they state can be reduced to physical laws. But since we have already seen that the major reductive program in psychology, philosophical behaviorism, faces serious difficulties. The conclusion seems to be, if we are to follow

Goodman's line of thinking, is that psychological predicates are not well-behaved. Goodman's position seems to lead us to the position of eliminative materialism as represented in figure 3.1.1.

However, Davidson has suggested that Goodman's account of projectibility overlooks the possibility that a predicate may be well-entrenched with respect to a family of predicates, but not all families of predicates are necessarily well-entrenched with respect to each other.[288] Thus the psychological predicates may be well-behaved with respect to each other, but still not well-entrenched with respect to the predicates of physics. But this raises a problem. If psychological predicates are well-behaved with respect to each other then, by Goodman's account of well-behavedness, they must have a past history of participation in lawlike statements; in other words, there must be psychological laws. But if there are psychological laws, and we are committed to materialism, then we have a mystery. How can two sets of "laws", i.e., psychology and physics, with no systematic relation to one another, both make predictions about the same material realm? It would seem that if psychology is indeed comprised of true lawlike statements, then it must be reducible to physics after all.

---

[288] This argument is untypical of Davidson in two respects. First, whereas Davidson usually stresses that our language and thought are holistic in nature, he is here suggesting that sets of predicates may be relatively insulated from one another. Secondly, whereas Davidson usually focuses on the behavior and characteristics of whole sentences in discussing philosophical problems, he is here basing an argument on the behavior and characteristics of parts of sentences, i.e., predicates. Davidson's holism and instrumental attitude toward sub-sentential entities will emerge more fully in Chapter Four.

In order to get out of this impasse I believe that we need an account of "pseudo-lawlikeness". That is, we need to show that psychological statements approach lawlikeness, but fall down in various ways. Their near lawlikeness confers a degree of Goodmanian projectibility upon psychological predicates, and warrants the Davidsonian claim that psychological predicates are well-entrenched with respect to each other. On the other hand, the systematic departures from true lawlikeness will dispel any worries about "miraculous" harmony between the laws of psychology and physics, and therefore we can comfortably maintain our position that psychology is not reducible to physics. However, the systematic departures cannot be too pervasive, for that would undermine our claim that psychological statements can approach lawlikeness in even a reduced sense.

Functionalism

A number of theorists have tried to maintain this difficult balancing act with a philosophy of psychology known as functionalism. Functionalism, in general, is the doctrine that psychological states are functional states, rather than a particular type of physical state. Fodor's particular version of functionalism, as we shall see, also incorporates the idea of mental representation as a key element. In order to make this clear the two key notions of functional system and mental representation must be explicated.

The notion of functional system is actually a familiar one. Some examples of functional systems are: the mousetrap, the airplane, the keyboard instrument, the microscope, the toaster, etc. Mousetraps, for example, form a class of objects and we recognize that the class of mousetraps, unlike the class

{ the moon, the number 7, Wagner's Die Walküre }

has some central organizing principle, such that all its members are naturally viewed as that type. But the organizing principle cannot be physical makeup, for mousetraps, as we know, can be made in all sorts of ways. Rather, the organizing principle of the class is a *function*, i.e., mouse-trapping.

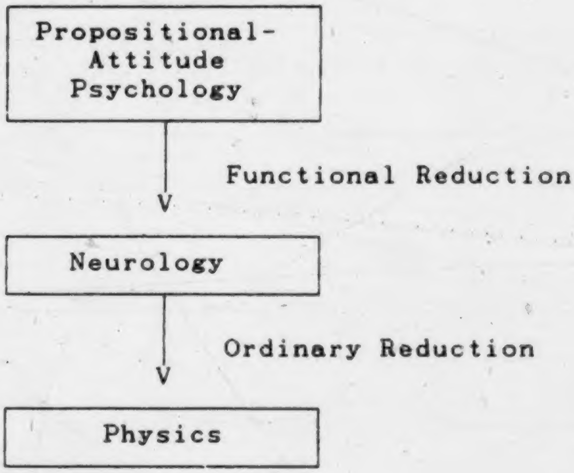
The concept of a functional system can be developed along obvious lines. We can speak of *functional components*, e.g., a particular variety of mousetrap is made up of a "base", a "trigger", a "spring", etc. In a complex functional system, e.g. a computer, there may be many layers of functional decomposition in which functional components are themselves broken down into functional components. We can also speak of *functional states*, e.g., a trigger-spring type of mousetrap may be in the "set" or the "sprung" state. A computer can have a enormously large number of functional states. Note that for any type of functional system, it will be the case that both its functional components and its functional states can be physically realized in a wide variety of ways.[289]

---

[289] For more on the notion of functional systems, see the essays in Part Three of N. Block, ed. *Readings in the Philosophy of Psychology*, Volume 1, (Harvard University Press, 1980).

The functionalist's views on the relation between psychology and physics are summed up in the following diagram.

Figure 3.4.1
Fodor on the Relation Between Psychology and Physics


Propositional-Attitude Psychology
Functional Reduction
Neurology
Ordinary Reduction
Physics

"Functional reduction" (not a standard term) is a relation that can hold between two theories. Functional reduction is weaker than ordinary reduction, but stronger than mere supervenience. I will not attempt to specify formal conditions on functional reduction, other than to note the following: Ordinary reduction, it will be recalled, is accomplished through bridge laws of the form:

For all x, Px iff Qx

where P is a predicate of the basic science (say statistical mechanics) and Q is a predicate of the science that is being reduced (say thermodynamics). But for functional reduction, the bridge laws will take the form:

(For all x, if P_1x then Qx) and
(For all x, if P_2x then Qx) and
:
:
(For all x, if P_nx then Qx)
:

That is, the bridge laws will take the form of an open-ended list of conditionals that express sufficient conditions, in terms of the basic science, for realizing the functional condition characterized by the predicate Q. Indeed, they are best not called "laws" at all, therefore they will be called "reducing formula" in what follows.

There are three important ways in which functional reduction is weaker than ordinary reduction.

1. Because of the open-ended character of the reducing formula, the complex physical predicate formed by the conjunct of these formula will not determine a physical natural kind, i.e., a predicate that plays a role in physical laws. Obviously, the class of physical objects constituting mousetraps will not appear as a natural kind in a physics textbook.

2. A functional reduction is always "defeasible" in the sense that if the system ceases to function in the intended sense even though the reducing formula are instantiated, we will not hesitate to modify the reducing formula. This is part of a more general attitude toward functional systems, whereby the achievement of the function, not the physical makeup of the system, is the constitutive principle. This quasi-normative principle is nicely brought out in the following passage by Marvin Minksy:

> When a particular machine is described to us, we do not first ask questions about its material construction. Given an engineering drawing, a circuit diagram, a patent description - something must first convince us that we understand how it works *in principle*. That is, we must see how it is "supposed" to work. We inquire only later whether this member will stand the stress, or whether that oscillator is stable under load, etc. But the *idea* of a machine usually centers around some *abstract* model or process.
>
> There is a curious contrast between this idea of a machine and the idea of a "theory". Consider some "theory" of physics, e.g., Newton's mechanics. This theory (or any other theory of physics) is supposed to be a generalization about some aspect of the behavior of objects in the physical world. *If the predictions that come from the theory are not confirmed, then (assuming that the experiment is impeccable) the theory is to be criticized and modified*, as was Newton's theory when the evidence for relativistic and quantum phenomena became conclusive. After all, there is only one such universe and it isn't the business of the physicist to censure *it*, much as he might like to.
>
> For machines, the situation is inverted! The abstract idea of a machine, e.g., an adding machine, is a *specification* for how a physical object *ought* to work. If the machine that I build wears out, I censure *it* and perhaps fix it. Just as in physics, the parts and states of the physical object are supposed to correspond to those of the abstract concept. But in contrast to the situation in physics, we criticize the *material* part of the system when the correspondence breaks down.[290]

3. Functional systems are typically not causally closed. I will not try to precisely define the notion of a "casually closed" system, but simply note that considered as mechanical systems, the solar system is relatively closed, whereas a fish in the sea is relatively not. My point is that any functional system will be a physical system characterized by some subset of the reducing formula, and that this physical system will

---

[290] M. Minsky, *Computation: Finite and Infinite Machines*, (Prentice-Hall, 1967), 5-6.

typically not be a causally closed system. For example, the system of financial transactions (throughout the world) can be viewed as a functional system, but the physical objects and events that instantiate it (i.e., pieces of paper and metal, states of computers and telecommunication systems, etc.) is certainly do not constitute a causally closed system. Consequently, objects and events that are not themselves financial transactions (e.g., a bank that burns down, a telecommunications satellite that fails) can affect financial transactions. This exposure to external disruptions is typical of functional systems.

Fodor's Assumptions

In a series of important books, Jerry Fodor has developed a complex defense of the functionalist approach to intentional psychology.[291] Some of Fodor's basic assumptions are the following:

1. Ordinary folk psychology, which explains and predicts people's actions as a quasi-rational function of their propositional attitudes, is basically on the right track. In other words, folk psychology is largely true.

2. An essential aspect of folk psychology is that its explanatory power derives, in part, from the "content" relations between the propositional attitudes held by the actor. The

---

[291] J. A. Fodor, Psychological Explanation, (Random House, 1968); J. A. Fodor, The Language of Thought, (Thomas Y. Crowell, 1975); J. A. Fodor, Representations, (The M.I.T. Press, 1981); and J. A. Fodor, Psychosemantics, (The M.I.T. Press, 1987).

content of a propositional attitude, such as a belief, a desire, etc. is the way the world has to be if that belief is to be true, that desire to be fulfilled, etc.[292] It is because of these content relations (because, for example, there is a content relation between the belief that if I pay the money, then I may own that stereo equipment and the desire that I own that stereo equipment) that the explanations of folk psychology have the explanatory force that they do.

3. Scientific psychology should be an elaboration of the basic explanatory model of folk psychology. That model is outlined in the following passage:

> o The agent finds himself in a certain situation (S).
>
> o The agent believes that a certain set of behavioral options (B_1, B_2, ...B_n) are available to him in S; i.e., given S, B_1 through B_n are the things that the agent believes he can do.
>
> o The probable consequence of performing each of B_1 through B_n are predicted; i.e., the agent computes a set of hypotheticals roughly of the form if B_i is performed in S, then, with a certain probability C_i. Which such hypotheticals are computed and which probabilities are assigned will, of course, depend on what the organism knows or believes about situations like S....
>
> o A preference ordering is assigned to the consequences.
>
> o The organism's choice of behavior is determined as a function of the preferences and the probabilities assigned.[293]

---

[292] This characterization of content is taken from J. A. Fodor, Psychosemantics, (The M.I.T. Press, 1987), 11.

[293] J. A. Fodor, The Language of Thought, (Thomas Y. Crowell, 1975), 28-29.

Another aspect of this framework not mentioned in this passage is the notion that beliefs and desires (preference orderings) also change over time. The framework, or "theory schema", as Fodor calls it, for scientific psychology must include all of these elements, and all these elements derive from folk psychology.

But why must we adopt the framework of folk psychology as a basis for scientific psychology? One of the central reasons that Fodor gives is that many of the classical problems of psychology can only be characterized from within this paradigm. Learning, for example, is a species of belief-formation, but the only way of distinguishing learning from other types of belief formation (e.g., taking a pill, getting bumped on the head, etc.), is to view learning as a computational relation between a set of propositions (or proposition-like entities) that represent (to the learner) the content of the sensory evidence before him, and a set of propositions (or proposition-like entities) that represent (to the learner) what is learned.[294] Almost all the concepts and interests of psychologists must, according to Fodor, be constructed in terms of propositional attitudes in this fashion.

---

[294] For a fuller development, see J. A. Fodor, "Computation and Reduction", in W. Savage, ed., Minnesota Studies in the Philosophy of Science, Volume 9: Perception and Cognition, (University of Minnesota Press, 1978). Reprinted in J. A. Fodor, Representations, (The M.I.T. Press, 1981), 152-156. For a similar approach to the difference between "rational" belief formation and belief formation in general, see D. Davidson, "Empirical Content", in E. LePore, Truth and Interpretation, (Basil Blackwell, 1986).

4. We should be realistic about propositional attitudes, according to Fodor. That is, people really do have propositional attitudes, and their attitudes have causal powers. The goal of a scientific psychology should be to describe the attitudes as they really are.

5. Psychology must be consistent with materialism, the thesis that all psychological events are physical events. Furthermore, there must be some way, short of classical reductionism, to show that psychology is in some sense integrated with, or at least not at odds with, natural science. (Note the location assigned to Fodor in figure 3.1.1.)

The Mid-Seventies Theory - The Strong RTM a.k.a. The LOT Theory

In the mid-seventies Fodor developed a version of psychological functionalism now known as the Strong Representational Theory of Mind (RTM), and which is also known as The Language of Thought (LOT) Theory.

This theory grew out of dissatisfaction with an early version of functionalism, known as Turing Machine (TM) functionalism. The idea of TM functionalism was that a psychological subject can be viewed as the instantiation of a Turing Machine (i.e., an abstract device characterized by inputs, outputs, and internal states such that the output and internal state at time t+1 are functions of the input and internal state at time t), and that psychological states are identified with internal states.[295] A particular psychological state, say the belief that snow is white, would be "reduced" to Turing Machine language by means of a statement such as the following:

> Subject S believes that snow is white if and only if subject S instantiates Turing Machine x, and subject S is currently "realizing" internal state C30A of Turing Machine x.

Fodor argued that there are a number of problems with this version of functionalism.[296] To mention just two of them: (1) TM functionalism holds that a psychological subject is in exactly one psychological state at a time (since Turing Machines are in exactly one internal state at a time). This leads to the unsatisfactory view that people can hold exactly one belief at a time. The TM functionalist could try to get around this objection by viewing psychological states as complexes of propositional attitudes, but this complicates the theory and diminishes its attraction. More importantly though, the model of psychological explanation reviewed above does not deal with complexes of propositional attitudes; rather, the psychological states that the model deals with are exactly one propositional attitude wide. (2) Turing Machine specifications are finite, and in particular, they have a finite number of internal states. Thus, according to TM functionalism, a psychological subject can have only a fixed and finite number of psychological states. This is unsatisfactory according to Fodor. The problem is not so much that TM functionalism limits the psychological subject to a finite number of states, it is the way that it does it. Psychological states are productive, i.e., new psychological states can be generated without limit. (This property naturally follows from the fact that psychological states are, according to the general position we are investigating, propositional attitudes.) Now it may be the case that because of physical finitude, all organisms are limited in the number of psychological states that they can undergo in a lifetime. But there is a difference between saying that one is limited to a finite number of psychological states, and saying that one is limited to a particular roster of states. TM functionalism entails the latter, which Fodor believes is surely false.

---

[295] For a fuller explanation of Turing Machines, see M. Minsky, Computation: Finite and Infinite Machines, (Prentice-Hall, 1967), chapters 6-10. For the classic exposition of TM functionalism, see H. Putnam, "Minds and Machines", in S. Hook, ed., Dimensions of Mind, (New York University Press, 1960); reprinted in H. Putnam, Mind, Language and Reality: Philosophical Papers, Volume 2, (Cambridge University Press, 1975).

[296] J. A. Fodor and N. Block, "What Psychological States are Not", Philosophical Review, 1972, 81. Reprinted in J. A. Fodor, Representations, (The M.I.T. Press, 1981), see esp. 87-97. Note that in this article the authors use the term 'Functional State Identity Theory (FSIT)' to refer to TM functionalism.

Fodor's new improved version of functionalism is based on the premise that there is a system of mental representation, a language of thought, if you like. This language of thought has a syntax (i.e., it is made up of "words" that can be concatenated together to form "sentences") and a semantics (i.e., "sentences" in the language of thought may be true or false depending on how the world is). Fodor claims that to have a propositional attitude is to be in a computational relation to a token of a sentence in the language of thought, e.g.,

> Subject S believes that snow is white if and only if subject S is in computational relation y to a token of sentence D42A of the language of thought.

Different propositional attitudes, i.e., believing, desiring, remembering, etc., will have different computational relations associated with them. Different "propositions", i.e., "snow is white", "grass is green", etc., will have different sentences in the language of thought associated with them.

Fodor also argues that it is at least conceivable that the language of thought is reducible to neurology. Asking himself the rhetorical question, "What would an acceptable reduction of psychology to a more fundamental science look like?", Fodor answers as follows:

> I suppose the answer goes like this: substantive reduction would at least require (1) that token computational processes turn out to be token neurological processes (storing a formula turns out to be a neurological process, etc.); (2) token internal representations turn out to be token neurological states (a token internal representation that translates 'elephants are gray' turns out to be some neurological configuration, in roughly, the way this sentence token is a configuration of ink marks on this page); and (3) canonical names of internal formula (viz., their structural descriptions) are specifiable in the vocabulary of neurology.
>
> I take it that (1) and (2) are just consequences of applying the usual ontological conditions upon reduction to the special case of psychological theories that acknowledge internal representations... It is (3) that does the work. In effect (3) requires that the canonical neurological description of a mental state (of a's) be of the form R_a, SD, [where SD is a structural description of a sentence of the language of thought] (and not, for example, of the form R_a, Alfred). So the question that has to be faced is: what would have to be the case for (3) to be satisfied? Heaven knows. I am unclear about how that question should be answered, but what I think it comes to is this: for psychology to be substantively reducible to neurology, it must turn out that the neurological entities constitute a code, and that the canonical

> neurological representation of such entities specifies the properties in virtue of which they constitute formulae in that code. Since the properties in virtue of which a formula belongs to a code are the ones in virtue of which it satisfies its structural description, and since the properties in virtue of which a formula satisfies its structural description, are the ones in virtue of which it has the content it has, we can summarize the whole business by saying that neurology will not rescue psychology unless neurological descriptions determine the content of internal formulae. (Compare the standard view, in which what specifies the content of a mental state is its canonical neurological representation _together with the relevant bridge laws_, and in which the specification is couched in the vocabulary of the reduced rather than the reducing science.)[297]

In summary, Fodor's position is that the propositional attitudes of folk psychology will have counterparts in a scientific psychology, viz., computational relations to tokens in the language of thought. The language of thought itself will be a neurological code. The (alleged) fact that the language of thought is neurologically based entails that both scientific psychology and folk psychology are integrated within the scientific world-view.

---

[297] J. A. Fodor, "Computation and Reduction", in W. Savage, ed., _Minnesota Studies in the Philosophy of Science, Volume 9: Perception and Cognition_, (University of Minnesota Press, 1978). Reprinted in J. A. Fodor, _Representations_, (The M.I.T. Press, 1981). Note that in earlier works, notably The Language of Thought, Fodor held the same position regarding the neurological basis for the language of thought, but tended to deny that this relation should be construed as a form of theoretical reduction. In the present work, he points out - appropriately, I think - that the language of thought hypothesis is a way of integrating psychology with a more fundamental science, albeit not exactly according to the model of classical theoretical reduction.

The "Meaning Ain't in the Head" Fiasco

About the same time that Fodor was developing this philosophy of psychology, Putnam had published his famous paper "The Meaning of 'Meaning'", in which he developed the Twin Earth examples and came to the conclusion that "meaning ain't in the head". To briefly recall Putnam's argument (cf. section 2.3), he claimed that the "classic" theory of meaning claims that:

(I) knowing the meaning of a term is just a matter of being in a certain psychological state, and

(II) the meaning (or intension) of a term determines its extension.

Point (II) can be generalized as follows:

(G) the meaning (or intension) of an expression determines its semantic properties (i.e., its denotation, extension, truth-conditions, etc., depending on whether it is a name, a predicate, a sentence, etc.).

Putnam's examples show that two people can be in an identical computational state (e.g., the state that Fodor would correlate with the psychological state of believing that this object that I am holding is a glass of water), but that those identical computational states do not determine the same semantic states of affairs (for one is holding a glass of H_2O and the other is holding a glass of XYZ). Putnam's examples seem to show that semantic properties are not supervenient on the computational or neurological states of a person.

Stephen Stitch has argued that Putnam's Twin Earth example, and other similar doppelganger examples that he and others have constructed, show that Fodor's program is in considerable trouble.[298] The reason, he says, is that scientific psychology, as well as any other field of science, must respect a "Principle of Autonomy". The autonomy principle is best explained by invoking the auxiliary notion of a replica. Suppose an exact molecule-for-molecule replica of an object O is constructed, and we call the replica O_R. Now an "autonomous description" of object O in a given setting S is any description that satisfies the following condition: if it applies to O in setting S, then it applies to O_R in setting S.[299] For example, if an exact replica is made of my eight-year-old car, then the description '... has some corrosion in the radiator' is autonomous (being true of both my car and the replica), but '... was purchased in 1986' is not autonomous (being true of my car, but not the replica). Now a condition on every nomological discipline is that it should deal with only autonomous descriptions, and psychology should be no exception. However, Putnam-type examples show that a person's intentional properties are typically non-autonomous. Stitch brings this out in the following passage:

> Suppose that in some distant corner of the universe there is a planet very much like our own. Indeed, it is so much like our own that there is a person who is my doppelganger. He is atom for atom identical with me and has led an entirely parallel life history. Like me, my doppelganger teaches in a philosophy department, and like me he has heard a number of lectures on the subject of proper names delivered by a man called 'Saul Kripke.' However, his planet is not a complete physical replica of mine. For the philosopher called 'Saul Kripke' on that planet,

---

[298] S. Stitch, "Autonomous Psychology and the Belief-Desire Thesis", The Monist, 1978, 61, 573-591.

[299] Adapted from S. Stitch, From Folk Psychology to Cognitive Science, (The M.I.T. Press, 1983), 167.

> though strikingly similar to the one called by the same name on our planet, was actually born in a state they call 'South Dakota', which is to the north of the state they call 'Nebraska'. By contrast, our Saul Kripke was born in Nebraska - our Nebraska, of course, not theirs. But for reasons which need not be gone into here, many people on this distant planet, including my doppelganger, hold a belief which they express by saying 'Saul Kripke was born in Nebraska.' Now I also hold a belief which I express by saying that 'Saul Kripke was born in Nebraska.' However, the belief I express with those words is very different from the belief that my doppelganger expresses using the same words, so different, in fact, that his belief is false while mine is true. Yet since we are doppelgangers the autonomy principle dictates that we instantiate all the same explanatory properties. Thus the belief property I instantiate in virtue of believing that Saul Kripke was born in Nebraska cannot be a property invoked in an explanatory psychological theory.[300]

In 1980 Fodor published a paper which granted that Putnam-type examples cause a problem for his Representational Theory of Mind.[301] Fodor outlined two broad strategies for dealing with the problem.

One strategy would be to drop the autonomy principle as a constraint on psychological theories. Fodor calls this approach "naturalism", in which "the recurrent theme ... is that psychology is a branch of biology, hence that one must view the organism as embedded in a physical environment."[302] A naturalistic psychology can individuate beliefs not only on the basis of what is in a believer's heads, but also on the basis of features of the believer's environments that are causally implicated in his beliefs. Consequently in a naturalistic psychology, Stitch's belief about Kripke can be distinguished from his doppelganger's belief about the other Kripke.

---

[300] S. Stitch, "Autonomous Psychology and the Belief-Desire Thesis", The Monist, 1978, 61, 581.

[301] J. A. Fodor, "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology", The Brain and Behavioral Sciences, 1980, 3. Reprinted in J. A. Fodor, Representations, (The M.I.T. Press, 1981).

[302] Ibid., 229.

Fodor rejected the naturalistic approach however. He argued that a naturalistic psychology would require laws containing predicates that apply not only to the internal states of psychological subjects, but also to all aspects of the environments of those subjects. In other words, a naturalistic psychologist is in Bloomfield's predicament: he must wait until all the other sciences are completed before psychology (semantics in Bloomfield's case) can begin. Fodor considers this an unacceptable constraint on psychology.

The other alternative is to retain the autonomy principle. (Any psychological theory that does so is called "rational psychology" by Fodor). However, the principle of autonomy requires that Fodor's original Representational Theory of Mind be modified somewhat. That theory was based on the idea that the explanatory power of psychological laws comes from the fact that they generalize over the "contents" of propositional attitudes. However, the Putnam-type examples show that contents are not an autonomous property of the psychological subject, and therefore psychological laws should not generalize over content. Fodor's solution (called the "Weak Representational Theory of Mind" by Stitch[303]), is to first of all recognize that the representational theory of mind has a "syntactic" side and a "semantic" side. By "syntactic" Fodor means that his theory postulates that there are entities (i.e., tokens of the language of thought) instantiated in the heads of psychological subjects, and furthermore, that his theory postulates that there are computational relations defined over those entities. But both the entities and the computational relations are formal, that is to say, syntactic in nature. And these syntactic properties are common to both Stitch and Stitch's doppelganger. That is, the syntactic properties are the psychological properties that we are left with when we properly observe the principle of autonomy. In Fodor's own terminology, syntactic properties are what we end up with when we practise "methodological solipsism", which is what we must do in order to avoid the pitfalls of naturalism.

---

[303] S. Stitch, From Folk Psychology to Cognitive Science, (The M.I.T. Press, 1983), chapter 9.

Psychological laws, then, should generalize over syntactic entities and properties. Fodor's original theory was that psychological laws should generalize over the semantic features of mental representations, that is to say, over contents. However, Fodor now realizes that Putnam-type examples point out that this would commit him to a violation of the principle of psychological autonomy. Furthermore, since Fodor strongly believes that a psychological theory should explain the mental causation of action he has another reason to couch psychological laws in terms of syntactic categories, for surely we would not want a theory that invokes semantic considerations (e.g., the actual circumstances of Kripke and Twin-Kripke's birth) in the explanation of action (e.g., of Stitch's and Stitch's doppelganger's actions). Clearly, these semantic considerations are irrelevant to the causal explanation of action.

However, this raises an important issue for which Fodor must provide an answer: what is the relation between the syntax in the head and the naturalistically based semantics of thought? Obviously Fodor must provide a satisfying answer to this question, for he views scientific psychology as a vindication of folk psychology[304], and the explanations of folk psychology certainly do appeal to the notion of content. The problem is that the syntactic notion of a mental state says that Stitch and Stitch's doppelganger are in the same mental state, whereas folk psychology (at least according to intuitions held by Putnam, Stitch, Fodor, and me) holds that they are in different psychological states. Fodor's answer is that they are different, but not that different. In the following passage he points out the difference:

> That taxonomy in respect of content is compatible with the formality condition [i.e., the principle of autonomy], plus or minus a bit, is perhaps the basic idea of modern cognitive theory... It's allowed that mental representations affect behavior in virtue of their content, but it's maintained that mental representations are distinct in content only if they are also distinct in form. The first clause is required to make it plausible that mental states are relations to mental representations and the second is required to make it plausible that mental processes are computations. (Computations just are processes in which representations have their causal consequences in virtue of their form.) By thus exploiting the notions of content and computation together, a cognitive theory seeks to connect the intensional properties mental states with their causal properties vis-à-vis behavior.

---

[304] J. A. Fodor, Psychosemantics, (The M.I.T. Press, 1987), 16.

> Which is, of course, exactly what a theory of mind ought to do.[305]

This is a very unsatisfactory response. Fodor has basically given some lip service to the fact that Putnam-type examples show that there is more to content than what is in the head. His fix is to claim that psychological laws should quantify over syntactic objects, rather than contents. But he suggests that things haven't really changed much because syntactic objects will stand in a one-to-one relation with belief contents, "plus or minus a bit."

In a book that deals with fundamental problems in the philosophy of psychology, Stitch has objected strongly to Fodor's facile response. After devoting two chapters to explicating the differences between the way our intuitions (i.e., folk psychology) differentiate beliefs compared to the way a "syntactic-causal" theory would individuate them, Stitch concludes:

> Granting that the narrow causal version of the mental sentence theory does not fully capture our folk psychological notion of belief, it is worth pondering just how badly it misses the mark. Fodor ... anticipated some slippage between the narrow causal standard of individuation and the content-based scheme of "aboriginal, uncorrupted, pre-theoretical intuition." However, on ~~his view~~, the two classification schemes will coincide "plus or minus a bit." The issue assumes considerable importance if, as Fodor maintains, the close (if imperfect) correspondence between the narrow causal (or "formal" taxonomy and the content-based taxonomy "is perhaps the basic idea of modern cognitive theory." ... [T]here is reason to doubt that cognitive science need be much

---

[305] J. A. Fodor, "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology", The Brain and Behavioral Sciences, 1980, 3. Reprinted in J. A. Fodor, Representations, (The M.I.T. Press, 1981), 240-241.

> concerned about the size of the gap of these two taxonomic schemes. But if Fodor is right on this point, then cognitive science is in deep trouble. For it is quite clear that by any reasonable measure the divide between folk taxonomy and the narrow causal taxonomy is enormous.[306]

Stitch's response to this "enormous" gap is to advocate a form of eliminative materialism that he calls the "Syntactic Theory of Mind" (STM). The STM is in agreement with Fodor's claim that cognitive science must be formulated in terms of causal relations that mirror formal relations between syntactic objects.

> The basic idea of the STM is that the cognitive states whose interaction is (in part) responsible for behavior can be systematically mapped to abstract syntactic objects in such a way that causal interactions among cognitive states, as well as causal links with stimuli and behavioral events, can be described in terms of the syntactic properties and relations of the abstract objects to which the cognitive states are mapped. More briefly, the idea is that causal relations among cognitive states mirror formal relations among syntactic objects. If this is right, then it will be natural to view cognitive state tokens as abstract syntactic objects.[307]

So far this is perfectly consistent with Fodor's theory. However, whereas Fodor wants to insist that these formal-causal patterns correlate, "plus or minus a bit", with the way that folk psychology individuates mental states in virtue of their content, the STM is "officially agnostic." Stitch claims:

> STM theories can do all the explanatory and predictive work of content-based theories, and they can do it better.... [The] inability of STM theories to capture the vague and observer-relative distinctions embedded in folk psychology is, I would argue, all to

---

[306] S. Stitch, From Folk Psychology to Cognitive Science, (The M.I.T. Press, 1983), 107.

[307] Ibid., 149.

> the good. There is no reason why a scientific psychology should respect the Protagorean parochialism of common sense.
>
> Fodor is clearly right that cognitive science began with the hope that content-based generalizations could be systematized and made more rigorous. As this effort has progressed, however, it has become increasingly clear that the most interesting and theoretically powerful generalizations are formal and syntactic ones which simply cannot be stated in the aboriginal language of content. If this is right, then surely the sane scientific strategy is to accept the STM paradigm and drop the attempt to characterize the interactions of mental states in terms of their content. It is, after all, a venerable tradition to kick away ladders once we have climbed them.[308]

Fodor on Contents as Functions

Let us review where we have been. We are investigating the prospects of a belief-desire psychology, for it has been argued that Whorf's Linguistic Relativity Hypothesis can only be made sense of in the context of such a psychology. I argued that propositional attitude psychology does not reduce to physics, at least not according to the classic model of theoretical reduction. But if psychology does not reduce to physics, then we have a puzzle, for Goodman has taught us that only some predicates, the "projectible" predicates, can participate in the statement of scientific laws. And it seems that what makes a predicate projectible is the extent to which it is "entrenched" in a system of other predicates, i.e., the extent to which the extension of that predicate can be put into some sort of systematic relation with other predicates. The "functionalist" doctrine appears to provide what we are looking for, for functionalism shows how psychological theories can be related to lower level theories without reducing to those theories. However, the original Turing-Machine version of functionalism is not viable, because, among other things, it does not properly account for the productivity of intentional states. A more viable version of functionalism is Fodor's Representational Theory of Mind, which views mental processes as computational relations between sentence like mental representations. A problem with this version of functionalism, however, arises out of Putnam's Twin Earth examples, for these examples show that representation is not merely a matter of what is in one's head; rather, it involves considerations about the subject's environment as well. Stitch has concluded that we should keep everything in Fodor's theory except the claim that the syntactic objects in the subject's head are representations of anything. That is, we should throw out semantic considerations, and concentrate on building a syntactic theory of mind. The old notions of "representation", "content", and "meaning" will end up on the scrap heap of "bad" concepts that also includes such notions as "the transmutation of matter", "demonic possession" and "phlogiston". (Note what this position means to the Linguistic Relativity Hypothesis. It means that we should throw it on the scrap heap too, for it is inexpressible without the notion of content.)

---

[308] Ibid., 182-183.

But Fodor has not been convinced. In a recent work he presents a more adequate defense of his claim that scientific psychology will be compatible with the notion of content as employed in folk psychology.[309] Fodor feels that theorists like Stitch have overreacted to the Putnam examples; they have assumed that we should throw the notion of content on the scrap heap because the traditional theory (that contents are in the head, and that contents determine extensions) doesn't work at all (i.e., there is no connection between content and extension). But all that the Putnam examples really show, according to Fodor, is that a content determines an extension, relative to a context.

> If you like, contents are functions from contexts and thought onto truth conditions...
>
> But now we have an extensional identity criterion for mental contents: Two thought contents are identical only if they effect the same mapping of thoughts and contexts onto truth conditions. Specifically, your thought is content-identical to mine only if in every context in which your thought has truth condition T, mine has truth condition T and vice versa.
>
> It's worth reemphasizing that, by this criterion, my Twin's 'water'-thoughts are intentionally identical to my water-thoughts; they have the same contents even though, since their contexts are de facto different, they differ, de facto in their truth conditions. In effect, what we have here is an extensional criterion for 'narrow' content. The 'broad content' of a thought, by contrast, is what you can semantically evaluate; it's what you get when you specify a narrow content and fix a context.[310]

As Fodor says, "if the worry about propositional attitudes is that Twin-Earth shows that contents don't determine extensions, the right thing to do is stop worrying."[311] In particular, Fodor concludes that his original notion that the syntactic objects in people's heads stand in a more-or-less one-to-one relation to contents is vindicated, provided, of course, that we understand contents as "narrow contents", i.e., as functions requiring a context "argument" in order to generate "broad content".

---

[309] J. A. Fodor, Psychosemantics, (The M.I.T. Press, 1987), chapter 2.

[310] Ibid., 47, 48.

[311] Ibid., 53.

3.5 AUTONOMOUS SOCIAL SCIENCE

Fodor's Representational Theory of Mind is, to my knowledge, the best case in the literature for the view that the propositional attitudes (and the folk psychology that is based on the propositional attitudes) can be accommodated within a scientific world view. If the Putnam examples were the only objection to Fodor's theory, I, for one, would be happy to entertain his position, for I believe that he has successfully shown that these examples do not have the dire consequences that they are often taken to have; I would accept that his approach will affect a reconciliation of the Geisteswissenschaften with the Naturwissenschaften, and that the Linguistic Relativity Hypothesis is, consequently, a normal scientific hypothesis to be evaluated according to normal scientific practises.

However, I do not hold this position, for I believe that there is another class of objections to Fodor's approach which he cannot overcome. Given that, in light of these objections, Fodor's "functionalist" attempt at reconciling the Geisteswissenschaften with the Naturwissenschaften fails, and given that "classical" theoretical reduction fails, and given that no other method of accomplishing this reconciliation seems available, I believe that the reconciliation is not possible. In other words, Brentano's thesis that the propositional attitudes are not explicable in terms of the natural sciences is true.

As reviewed earlier in this chapter, if Brentano's thesis is true, then we must decide what to do about the propositional attitudes. Either we must consign them to the scrap heap, as Stitch does, or we can study them through an autonomous social science. I believe that the latter course is correct.

The purpose of this section is to spell out the argument outlined in the last three paragraphs. More precisely, I have two objectives in this section. The first is to show that Fodor's functionalism falls apart because a number of reasons other than the context dependence of propositional attitudes that is brought out by the Twin Earth examples. The second is to argue that we should prefer the idea of an autonomous social science to eliminative materialism.

Reasons for Rejecting Fodor's Functionalism

Fodor's functionalism appears to be our best bet for reconciling propositional attitude psychology with the natural sciences. However, I will give two reasons why I don't think it will work.

In order to forestall any possible misunderstanding, note that I am not claiming that Fodor's idea of a language of thought is a bad one. What I am questioning is whether this alleged language of thought has a close connection with the type of propositional attitude explanations that we find in folk psychology and the social sciences. Recall that Fodor and Stitch agree, as a working hypothesis, that there is a formal language of thought, and that cognitive processes are viewed as computational relations over neurologically instantiated tokens in the language of thought. But Fodor goes on, and Stitch does not, to say that the language of thought is closely tied to our ordinary propositional attitude psychology. It is this last step of Fodor's that I want to question.[312]

There are, I believe, two good reasons to reject Fodor's claim that the propositional attitudes that we employ in folk psychology and social science are correlated fairly closely with the syntactic objects postulated in the language of thought theory.

1. Attitudes without Internal Correlates

First of all, it seems quite clear that one can attribute propositional attitudes to a psychological subject even in cases where one knows quite well that the subject has no internal correlates to the propositional attitudes. Dennett makes this point in the following passage:

---

[312] Although I am not objecting to the hypothesis that Fodor and Stitch share (i.e., the Syntactic Theory of Mind), I am not committing myself to it either. As noted in the discussion of Minsky's frame theory in section 3.2, a concern with the syntactic theory of mind is that sentence-like objects require a lot of computational processing in order to accomplish cognitive tasks. For this reason, a number of cognitive scientists have moved away from sentence-like objects in their computational models of cognitive processes. For further discussion, see D. C. Dennett, "The Language of Thought Reconsidered", in The Intentional Stance, (The M.I.T. Press, 1987).

> In a recent conversation with the designer of a chess-playing program I heard the following criticism of a rival program: "It thinks it should get its queen out early." This ascribes a propositional attitude to the program in a very useful and predictive way, for as the designer went on to say, one can usually count on chasing that queen around the board. But for all the many levels of explicit representation to be found in that program, nowhere is anything roughly synonymous with "I should get my queen out early" explicitly tokened. The level of analysis to which the designer's remark belongs describes features of the program that are, in an entirely innocent way, emergent properties of the computational processes that have "engineering reality". I see no reason to believe that the relation between belief talk and psychological-process talk will be any more direct.[313]

Fodor has responded to this criticism by granting that Dennett's complaint has some force, but not in a way that causes problems for his theory. He admits, and in fact insists, that not every rule that an intelligent system obeys will be internally represented in that system. A system that only represents will not do anything; therefore some of the "rules" are actually abstract descriptions of the way the system works, rather than what it is representing. On the other hand, some rules may be internally represented (like a program that is stored in memory in a Von Neumann type computer), so long as there are other "rules", not internally represented, that will cause the represented rules to be "executed." So Fodor's diagnosis of Dennett's example is that Dennett happened to focus on a "rule" that is not internally represented, but instead is an abstract description of the way the system works. But, as indicated toward the end of the following passage, Fodor holds that Dennett's own example probably does support his representational theory for most of the propositional attitude ascriptions that we are likely to make about that system:

---

[313] D. C. Dennett, "A Cure For the Common Code", in Brainstorms, (The M.I.T. Press, 1978), 107.

> According to the Representational Theory of Mind, programs - corresponding to the 'laws of thought' - may be explicitly represented; but 'data structures' - corresponding to the contents of thought - have to be.
>
> In Dennett's chess case, the rule 'get it out early' may or may not be expressed by a 'mental' (/program language) symbol. That depends on just how the machine works; specifically, on whether consulting the rule is a step in the machine's operations. I take it that in the machine that Dennett has in mind, it isn't; entertaining the thought 'Better get the queen out early' never constitutes an episode in the mental life of that machine... By contrast, the representations of the board - of actual or possible states of play - over which the machines computations are defined must be explicit, precisely because the machine's computations are defined over them. These computations constitute the machine's 'mental processes,' so either they are causal sequences of explicit representations, or the representational theory of chess playing is simply false for that machine.[314]

Dennett's point was that Fodor seems to be on the wrong track, because we make intentional attributions even in cases where we know full well there is no internal correlate to the attribution. Fodor's response is that the fact internal correlates are sometimes "missing" should be no surprise, because a system that had explicit representations corresponding to all of its intentional characteristics would not do anything, it would just sit there and represent. However, Fodor goes on, even in the chess-playing program case that Dennett raises, there will likely be a significant amount of correlation between the contents of the propositional attitudes we ascribe to the chess playing program and the explicit internal representations that have been provided for that program. Consequently, even in the case that Dennett has chosen, the correlations between intentional ascription and internal representation are far more significant than the occasional rule that is not explicitly represented. For example, when we say that the computer doesn't want to move its queen's knight into QB_3, this intentional ascription will correspond to a computational process over the relevant data structure.

---

[314] J. A. Fodor, Psychosemantics, (The M.I.T. Press, 1987), 1988, 25.

However, I do not see how any of this takes the sting out of Dennett's criticism. For assume that we have a chess-playing program with the kind of data structures, and therefore the kind of intentional-computational correlations, just noted. But for every program that is rich in data structures in this way, there will be an indefinite number of other programs which compute the same input-output function, but that have data structures that make no sense in terms of our ordinary theories of chess, or that do not have data structures at all. (Think of explicit representation as the instruction tape of a Universal Turing Machine. Now, for every pair consisting of a Universal Turing Machine and an instruction tape, there is a special purpose Turing Machine with no instruction tape that will behave exactly the same way. In fact there are an indefinite number of such special purpose Turing Machines.) These alternate programs are nothing but implicit rules; they have no data structures. But there is no doubt that if we were viewing only the external behavior of these we would make exactly the same intentional attributions that we would for the program that does utilize data structures - provided, of course, that we were not prejudiced by knowledge of how the program operates.

Consequently Dennett's point stands. We may grant that cognitive processes are computational processes, and we may even grant that they are computational processes defined over data structures, but so granting gives us no reason to assume that the data structures - if indeed, there are any - stand in any close relation to the contents of intentional attributions.

Of course the foregoing does not constitute a refutation of Fodor's claim that there is a correlation. It just points out that intentional attributions can function quite nicely without the correlation. Furthermore, it shifts the burden of proof to Fodor, for it is now up to him to explain why there needs to be a correlation.

2. Holism

Now for a positive argument that the correlation breaks down. If Fodor's correlation thesis is true, then for every belief that a person might have, there is some mental sentence, such that being in a certain computational relation to a token of that mental sentence is a sufficient condition for having that belief. Thus, to believe that $E = mc^2$ is to be in that particular computational relation. So Fodor must hold that Albert Einstein's coming to believe that $E = mc^2$ was coincident with his coming to be in that particular computational relation.

Now consider two other cases. (1) Back in 30,000 B.C. a man is hit on the head by a coconut and as a result ends up in the same computational state that Einstein was in when he first came to believe that $E = mc^2$. (2) A six year old child is given a short lecture on the meaning of 'E', '=', 'm', 'c', and the role of '2' as an exponent. He is then informed that it is a law of physics that $E = mc^2$.

In the first case our intuitive judgment is that our ancestor did not have the same belief that Einstein had, even though they were, by hypothesis, in the same computational state. My intuitive rationale for denying that the earlier man has the same belief as Einstein is that a belief cannot stand alone, as it were. Einstein's belief that $E = mc^2$ depends on the fact that Einstein has many other beliefs that support and inform that belief. My intuitions are shared by many other writers, including Stephen Stitch, from whom this example was adapted.[315]

In the second case we might be willing to allow that the child has a similar belief to Einstein's, but we want to deny that he has the identical belief. Again, the reason is that Einstein's belief is supported and informed by many other beliefs that the child does not have. But in terms of the computational notion of thought advanced by Fodor, is there really a difference between the child and Einstein? On that theory, all it takes to be in a belief state is to be in a certain computational relation to a token of a single sentence in the language of thought. I can think of no reason why Fodor would deny that the child and Einstein are not computationally related to the same mental sentence, and therefore he must be committed to the view that they have exactly the same belief. But this conflicts with our intuition that they have distinct beliefs.

---

[315] S. Stitch, From Folk Psychology to Cognitive Science, (The M.I.T. Press, 1983), 53-60.

These examples show that our pre-theoretical intuitions about belief individuation are that the content of a belief is determined, in part, by the other beliefs that one holds. This doctrine is called "holism", and it causes a problem for Fodor's program as he himself indicates in the following passage:

> Presumably an event (e.g., the production of behavior by some organism) would fall within the domain of such a psychology in virtue of instantiating one of its generalizations. And presumably such generalizations would apply to an organism at a time in virtue of the intentional state(s) that the organism is in at the time. THe way it ought to go is that the theory says things like: 'From any organism x that believes such and such and desires so and so, you get behaviors of the type ... blah.' You can, therefore, use the theory to predict that *this* organism x will give behavior of the type ... blah *if* you can identify this x as believing that such and such and desiring that so and so. This is just a long form of the truism that one way that intentional psychologies achieve generality is by *quantifying over all that organisms that are in a specified intentional state*.
>
> But now, if - as surely is the case - people quite generally differ in their estimates of epistemic relevance [i.e., Einstein and the child have very different views of the relation between $E = mc^2$ and other laws of physics], and if we follow Meaning Holism and individuate intentional states by the *totality* of their epistemic liaisons, it's going to turn out de facto that no two people (for that matter, no two time slices of the same person) ever *are* in the same intentional state. (Except, maybe, by accident.) So no two people will ever get subsumed by the same intentional generalizations. So intentional generalizations won't, in fact, succeed in

generalizing. So there's no hope for an intentional psychology.[316]

Fodor, of course, refuses to take this lying down, and consequently he devotes considerable effort to showing that the doctrine of Holism is on the wrong track.[317] His strategy is to assume that Holism is a philosophical theory, rather than something that is more or less given by pre-theoretical intuition, and therefore the burden of proof is on the advocates of Holism. He then considers three philosophical arguments for Holism, and while not definitively refuting them, he believes he has cast enough doubt on them to tip the balance in his favor.

However, I disagree with Fodor's assessment that Holism is a sophisticated philosophical doctrine, far removed from pre-theoretical intuition. I believe that it is a far more natural doctrine than he supposes, as evidenced by my reader's agreement (assuming there is agreement) with my intuitions regarding the $E = mc^2$ beliefs.

In any case, I believe that there is a good philosophical argument for Holism. One of the three arguments that Fodor rejects he calls the argument "from Confirmation Holism to Meaning Holism."[318] The gist of this argument is as follows: In his classic paper "Two Dogmas of Empiricism", Quine criticized the idea that scientific knowledge is confirmed one sentence at a time. Rather, "the unit of empirical significance is the whole

[316] J. A. Fodor, op. cit., 57.

[317] Ibid., chapter 3.

[318] Ibid., 62-67.

of science."[319] Quine argued that any given sentence may be held true no matter what evidence comes up, provided that we are willing to make enough changes in the truth values of other sentences in the theory. Fodor calls this doctrine "confirmation holism", and apparently he has no objections to it. What he does object to is the idea that confirmation holism entails what he calls "semantic holism", or the holism of belief attribution. Fodor's reasoning goes as follows:

> [Quine] rejects local semantic connections because they would imply that there are unrevisable statements. And he rejects the claim that there are unrevisable statements because it is false to scientific practice. In short, Quine's tactic is to infer Confirmation Holism from the refutation of semantic localism, and not the other way round. If, however, that is how the argument goes, then a Quinean cannot offer Confirmation Holism as an argument for Meaning Holism. That would be to argue backward.[320]

I reject this interpretation of Quine. What Quine is doing in the last two sections of "Two Dogmas" is recommending the replacement of foundationalist theories of knowledge with a holistic theory of knowledge. How very much against the spirit of Quine's article to read him as arguing from certain premises in a certain direction to certain conclusions. Rather he is presenting an account of how science works, and through the course of the article he draws out various aspects of that account. And I believe that the account of science that Quine presents is one in which the identity conditions of individual scientific beliefs (insofar as they can be drawn at all) are clearly dependent on their inferential relations to the other members of "the web of belief." So in opposition to Fodor I believe that confirmation holism does support meaning holism, not in the sense that confirmation holism entails meaning holism, but in the sense that both confirmation holism and meaning holism are aspects of the non-foundationalist theory of knowledge that Quine did so much to promote.

---

[319] W. V. O. Quine, "Two Dogmas of Empiricism", in From a Logical Point of View, 2nd ed., (Harper and Row, 1961), 42.

[320] J. A. Fodor, op. cit., 65.

So both pre-theoretical intuitions and epistemological arguments support the view that the individuation of a belief depends on the epistemic relations between that belief and others that one holds. This holistic view of beliefs is incompatible with Fodor's thesis that the propositional attitudes are correlated with very specific and narrowly defined internal computational processes.

Eliminative Materialism or Autonomous Social Science?

Fodor's functionalism was our last chance to avoid Brentano's irreducibility thesis; to try to find a home in the natural sciences for the propositional attitudes. This is not to deny that a cognitive science may develop in which the basic insight is that cognitive processes are computational processes defined over neurological entities that are construed to have a sentence-like syntax. It is just to deny that a psychology of this form will have much to do with propositional attitudes.

The question we are now facing is what should we do about the propositional attitudes? Should we turn our back on them because we can't fit them into our scientific world view, or should we make a case for an autonomous social science that is based on propositional attitudes? (This is the second question from figure 3.1.1.)

The first thing to note is that all parties in the debate agree that giving up the propositional attitudes would be a truly staggering historical event. Fodor writes:

> If commonsense intentional psychology really were to collapse, that would be, beyond comparison, the single greatest intellectual catastrophe in the history of our species; if we're that wrong about the mind, then that's the wrongest we've been ever been about anything. The collapse of the supernatural, for example, didn't compare; theism never came close to being as intimately involved in our thought and our practice - especially our practice - as belief/desire explanation is. Nothing except, perhaps, our commonsense physics - our intuitive commitment to a world of observer-independent, middle-sized objects - comes as near our cognitive core as intentional explanation does. We'll be in deep, deep trouble if we have to give it up.[321]

Similarly, Stitch - who is not adverse to the idea of giving up on the propositional attitudes - has pointed out the profound implications that such a move would have for the social sciences:

> Economics, political science, sociology, and anthropology are up to their ears in the intentional idiom that is the hallmark of folk psychology. If all talk of beliefs, desires, expectations, preferences, fears, suspicions, plans, and the like were banished from the social sciences, those disciplines as we know them today would disappear. We simply have no way of recounting our knowledge of social, political, and economic processes without invoking the intentional language of folk psychology... If, as Laudan urges, we choose between theories largely on the basis of the problems they solve, then, for all their evident difficulties, the social sciences will be around for the foreseeable future. For there are simply no serious competing theories which address problems in

---

[321] Ibid., xii.

> the social domain and which do not invoke the intentional concepts of folk psychology.[322]

It is obvious, then, that intentionally based social sciences will not disappear tomorrow. But what about the long term? If our previous arguments are sound, and intentional psychology cannot be reconciled with natural science, might belief/desire talk not slowly whither away as a new non-semantic cognitive science develops? Obviously one's answer to this question will be speculative in nature, and it may reveal more about the one's taste and temperament than anything else. In any case, I side with Daniel Dennett in predicting that the intentional disciplines will not whither away.[323]

Dennett believes that intentional ascriptions are normatively based, and that they are instrumental. In saying that they are normatively based, he is saying that intentional attributions must conform to an overall norm of rationality. More will be said about the particular norm of rationality in the next chapter, but for now just note that a normative attribution (of any type) is never really falsified when things don't turn out according to the attribution. Rather (as Minsky pointed out in his comparison of abstract machines and theories), we tend to censure the concrete reality as an "imperfect" realization of the normative standard. Many of our concepts of social science are exactly like this; for example, the perfectly rational consumer, the authoritarian personality, and the many other "ideal types" that we meet in social science.

---

[322] S. Stitch, op. cit., 213-214.

[323] Dennett's position is presented in the essays collected in his Brainstorms, (The M.I.T. Press, 1978), and The Intentional Stance, (The M.I.T. Press, 1987).

Dennett's instrumentalist attitude means that for him, propositional attitude psychology is a useful predictive calculus, but not something that is literally true or false. Furthermore, Dennett's instrumentalism means that he is not particularly worried if many of the terms used in folk psychology turn out not to have a reference.

Dennett's normative-instrumentalist view of propositional attitude psychology exempts it from the standards usually applied to scientific theories. Provided that the propositional attitudes continue to do what they do for us in everyday life and in social science, they are guaranteed a place in our conceptual scheme.

Stitch has presented two arguments against Dennett's view that propositional attitude psychology is an instrumental system. One argument is that only real entities, and not fictional instrumental entities, can have causal properties. But folk psychology entails that beliefs and desires have causal properties. Therefore folk psychology endorses a realist approach to mental entities, not the instrumentalist approach that Dennett claims it endorses.

I think there is a way out of this objection. It would be to define a relation between instrumental entities and underlying physical entities that is weaker than supervenience, but which establishes at least a "temporary identification" of those entities, such that the instrumental entities inherit the causal properties of the physical entities. I will not attempt to work out the details, but I see no reason why this approach would not take care of Stitch's concerns about causality.

Stitch's other objection is that instrumentalism is too liberal. According to instrumentalism, as long as an organism or artifact behaved in a suitably complex way, we would be willing to assign it propositional attitudes.

> But surely there are things we might find out which would convince us that a "person" who behaved normally enough did not really have beliefs. Ned Block has conjured the case of a chess-playing computer which, though it plays a decent game of chess, has no internal representation of rules or goals or strategy. Rather, this computer's memory contains an enormous multiply branching tree representation of every possible chess game up to, say, 100 moves in length. At each point in the game, the computer simply plays the appropriate prerecorded move. Watching the play of such a machine, we might be tempted at some point to say, "It believes I am going to attack with my queen." But on learning just how the machine works, we are much less inclined to say this. Analogously, if we were to run across (what appeared to be) a person whose conversations, chess playing, and other behaviors were controlled by an enormous, pre-programmed branching list of what to do when, I think our intuition would rebel at saying that the "person" believed that I was about to attack with my queen - or indeed that he believed anything else! Entities with innards like that don't have beliefs.
>
> ...But once this has been granted, it can no longer seriously be maintained that beliefs and desires are instrumentalistic.[324]

I think that Stitch is not being careful enough with this example. Consider this scenario: Suppose that in fifty years cognitive science (a nice clean "syntactic" cognitive science) determines that we are all constituted like Stitch's "person", i.e., we are all controlled "by an enormous, pre-programmed branching list of what to do when." Would we then conclude that none of us never had any beliefs? I think not. Instead we would express surprise that the having of beliefs turned out to be like that. Of course we do not expect that we are constituted in this way (i.e., the way insects are). Instead, we expect that we are inherently flexible and adaptable, and the ideas of flexibility and adaptability are associated with the idea of "intelligence". Stitch's "person", we all feel, is not very intelligent, at least not deep down. So we all believe that there is a contrast between ourselves and Stitch's "person"; we are intelligent and he is not. In fact we would probably be so impressed with the contrast that we may, as Stitch suggests, be reluctant to attribute beliefs to him. But again, the real reason for the feeling of contrast is simply our feeling that he doesn't have beliefs the way that we do. If we were to find out that, contrary to our expectations, we are constituted the same way as Stitch's "person", I do not believe that we would give up the concept of belief. If push came to shove, we would give up on the idea of "intelligence" long before we would give up on the idea of belief.

---

[324] S. Stitch, op. cit., 244-245.

To summarize, this chapter began with the idea that thought, and therefore Whorf's Linguistic Relativity Hypothesis, should be understood within the context of a propositional attitude psychology. It was argued that propositional attitude psychology could not be reconciled with natural science. Natural science will include a cognitive science, but it will be a "syntactic" science, stripped of all "semantic" pretensions, and therefore will have nothing to do with the Linguistic Relativity Hypothesis. Propositional attitude psychology, on the other hand, will be the foundation of the normatively and instrumentally constituted *Geisteswissenschaften*, which are "autonomous" from the natural sciences. The Linguistic Relativity Hypothesis is a hypothesis that belongs to the *Geisteswissenschaften*.

CHAPTER FOUR - RELATIVITY

4.1 INTRODUCTION

In Chapters Two and Three it was argued that language and thought are best conceptualized from within the framework of a general theory of social action, and because of this, language and thought belong to a field of study that is distinct from the natural sciences. It is time now to turn back to the Linguistic Relativity Hypothesis. My strategy for this chapter will be to focus on the problem of "interpretation", by which I mean the activity of determining the language (i.e, a specific truth-conditional semantic theory) and the thought (i.e., the specific beliefs, desires, etc.) of the members of a particular speech-community.

The previous two chapters developed a specific approach to language and thought. We now require a method for interpretation that is consistent with that approach. This method will be unique to the theory of social action, i.e., it will be distinct from the methods employed in the natural sciences. Now upon examination, we might find that our method for interpretation is inconsistent with the Linguistic Relativity Hypothesis. This could happen for any of the following four reasons.

1. The method of interpretation might entail that the attribution of meaning (language) and beliefs (thought) is a single enterprise, which cannot be distinguished as two distinct phases, one corresponding to the determination of language and the other to the determination of thought. In its strongest form the Linguistic Relativity Hypothesis postulates a causal relation between language and thought. But if these two things cannot be cleanly separated in interpretation then how can we justifiably distinguish them as separate entities (processes or mechanisms within the language user) that are capable of participating in a causal relation?

2. The method of interpretation might entail that radical differences in the thought of different speech-communities are not possible. That is, the method might be such that if the members of another speech-community can be interpreted at all, then the thoughts attributed to them will necessarily be fairly similar to our own. The incommensurability (or near-incommensurability) hypothesized by Whorf will turn out to be an impossibility.

3. The previous point notwithstanding, any method of interpretation must countenance the fact that there will be

differences of opinion that the method must accurately portray. But what if the very nature of the method demonstrates that we should expect to see just as wide a variation of conceptual scheme and opinion within a speech-community as between speech-communities? Then we would have to give up the Linguistic Relativity Hypothesis, for then it would be clear that language is no more significant a belief-determiner than other factors, if it is a belief-determiner at all.

4. Finally, the method may turn out to be such that interpretation is seen to depart significantly from the standards of empirical measurement that we expect in the sciences. In particular, it may turn out that interpretation lacks objectivity in that normative elements come into play. Furthermore, interpretation may lack determinacy in that the method permits many alternative albeit significantly different interpretations of a particular language-user. Perhaps this point in itself is not sufficient to topple the Linguistic Relativity Hypothesis, but in combination with any of the previous points, it speaks strongly against it.

In fact, I think that all four of these points are not just possibilities; the complications raised actually do exist, and consequently the Linguistic Relativity Hypothesis must be abandoned. The goal of this chapter will be to flesh this argument out.

In section 4.2 Attitudes, Meaning and Physics, I will argue that it will not be possible to develop a method that solves for language and thought independently. Instead, language and thought must necessarily be attributed in a joint methodological enterprise. This section will demonstrate point 1 above, and go some way toward demonstrating point 4.

In section 4.3 Radical Interpretation, I will present the views of some authors as to how this joint interpretation is to take place. The main point of this section will be to demonstrate that the method is subject to a number of necessary constraints that set a limit to what the output of any interpretation will be. This section is largely intended as a prerequisite for the next three, but in it I also raise the concerns expressed in point 4.

In section 4.4 Davidson on Conceptual Relativity, I examine Donald Davidson's views on the idea of incommensurable conceptual schemes. Although not entirely agreeing with Davidson's detailed arguments, I concur with him that methodological considerations force us to rule out the idea of radical conceptual differences. This establishes point 2.

In section 4.5 Relativity of Reference, I examine the views of Bruce Aune who argues that Davidson does not give enough latitude for conceptual variation. Aune holds that if we attend to the referential structures in language, we can attribute a greater range of conceptual variation. However, it is a consequence of Aune's view that conceptual variation can occur to the same degree within a speech-community as between speech-communities. This provides evidence for point 3.

In section 4.6 Relativity of Reasoning, I examine the views of Ian Hacking who also argues that Davidson does not give enough latitude for conceptual variation. Hacking holds that the patterns of reasoning can vary between groups of people, though this can happen to the same degree within a speech-community as between them. This provides further evidence for point 3.

In section 4.7 Conclusion, I review the case against the Linguistic Relativity Hypothesis.

4.2 ATTITUDES, MEANING, AND PHYSICS

David Lewis is a philosopher who subscribes to the view that natural languages are best understood by means of a truth-conditional semantic theory and that thought is best understood as the workings of a system of propositional attitudes. In his paper "Radical Interpretation", Lewis states the methodological problem that must be faced by anyone who accepts this picture of language and thought.

> Imagine that we have undertaken the task of coming to know Karl as a person. We would like to know what he believes, what he desires, what he means, and anything else about him that can be explained in terms of these things. We seek a two-fold interpretation: of Karl's language and of Karl himself. And we want to know his beliefs and desires in two different ways. We want to know their content as Karl would express it in his own language, and also as we would express it in our language....
>
> Imagine also that we must start from scratch. At the outset we know nothing about Karl's beliefs, desires and meanings. Whatever we may know about persons in general, our knowledge of Karl in particular is limited to our knowledge of him as a physical system. But at least we have plenty of that knowledge - in fact we have all we could possibly use. Now, how

can we get from that knowledge to the knowledge we want?

I can diagram the problem of radical interpretation as follows. Given P, the facts about Karl as a physical system, solve for the rest.

P Karl as a physical system

Ao Karl's attitudes: beliefs and desires as expressed in our language

Ak Karl's attitudes: beliefs and desires as expressed in Karl's language

M Karl's meanings: truth conditions of his full sentences

(and denotations, etc. of constituents of sentences)

[325]

What we have is a single phenomenon, Karl and his activities, that can be described in several ways. In Lewis' formulation of the problem of radical interpretation, one description, the physical, is given to us. The problem is to relate this physical description to the other descriptions that reveal Karl as a person. If any two theories or descriptive systems both apply to a single empirical domain, a natural question is: what is the nature of the relation between the two theories or descriptive systems? The activities in a test tube are described by a physicist and a chemist. What is the relation between the physicist's language and the chemist's? A man signs a cheque to give a large sum of money to charity. This can be described in physical terms (at least in principle), in biological terms, in psychological terms, and in moral terms. But what is the relation between these alternate descriptions?

---

[325] D. Lewis, "Radical Interpretation", Synthése, 1974, 23, 331-344. Reprinted in D. Lewis, Philosophical Papers: Volume I, (Oxford University Press, 1983), 108-109.

There are three ways that a pair of alternate descriptions can be related to each other:

1. Reduction

The idea here is that of the two descriptions, one is more "fundamental", with physics being more fundamental than chemistry, chemistry more fundamental than biology, etc. Each predicate of the less fundamental discipline has an extension that can be stated in a fairly compact open sentence using the vocabulary of the fundamental discipline. A statement of these extensional equivalences is sometimes called a bridge law. Two descriptive systems that are reductively related to one another are sometimes said to be in a "type-type" relation, for any "type" of entity, event, process, condition, etc. that can be described by the less fundamental discipline can also be described as a "type" of entity, event, process, condition, etc. from the perspective of the fundamental discipline. Finally, it should be mentioned that in some quarters there is a prejudice that if a descriptive system cannot be ultimately reduced to physics (perhaps by many intermediate reductions) then that descriptive system should not be allowed to participate in the formulation of our scientific knowledge.

2. Supervenience

Another possible situation is that there are two descriptions of a phenomenon, and once again, one description is more fundamental than the other. However, in this case although each individual entity, event, process, condition, etc. that is describable in the vocabulary the less fundamental discipline is also describable in the more fundamental discipline, it is not true that each type of entity, event, process, condition, etc. that can be characterized in the less fundamental discipline can also be characterized as a type in the fundamental discipline. An example might be helpful here.

> Suppose a hurricane, which is reported on page 5 of Tuesday's _Times_, causes a catastrophe, which is reported on page 13 of Wednesday's _Tribune_. Then the event reported on page 5 of Tuesday's _Times_ caused the event reported on page 13 of Wednesday's _Tribune_.[326]

This example shows that the events in question can be described as newspaper-reported-events. But it should be clear that the "discipline" of events-that-are-described-in-newspapers is not going to be type-type reducible to a more fundamental discipline, say meteorology, even though every individual token of a newspaper-described-event may be describable at that lower level.

3. Instrumental Alternatives

A third possibility is that there are two descriptive systems that are not even related at a token-token level. That is, the less fundamental discipline may postulate entities, events, processes, conditions, etc. that cannot be described in the more fundamental discipline. All the same, the less fundamental discipline may provide limited predictive powers that warrant its consideration as an instrumental alternative to the more predictively accurate fundamental discipline. Daniel Dennett suggests that this happens when a person plays chess with a computer. Typically the person will try to deal with the computer in psychological terms (as an "intentional system", to use Dennett's phrase), and will try to understand the play of the computer using attributions such as "It hasn't noticed what I am trying to do with my knight", or more general observations such as "It is very good at the opening game." Some of these statements may turn out to be supervenient on a description of the computer's chess-playing program, and indeed, it may even be possible to reduce some of the attributions. But as Dennett points out, many if not most of our psychological attributions will have no direct counterpart in the language of the computer program that instantiates the chess-playing abilities, either at the type-type or the token-token level. A specific belief, desire or pattern of reasoning postulated at the intentional level may have no counterpart at the program level because the programmer has used some obscure algorithm that is both different from and far more finely-structured than what we find in folk psychology.

---

[326] D. Davidson, "Actions, Reasons, and Causes", _Journal of Philosophy_, 1963, _60_. Reprinted in D. Davidson, _Essays on Actions and Events_, (Oxford University Press, 1980), 17.

Having explored these three ways in which descriptive systems can be related (or not related) let us return to Lewis' problem of radical interpretation. As formulated by Lewis, the problem is somehow to "derive" a description of Karl's attitudes and meanings from a description of him as a physical system. That is, we have three descriptive systems or "disciplines" that have to somehow be connected.

A Karl's attitudes

M Karl's meanings

P Karl as a physical system

(I have ignored for now Lewis' distinction between Karl's attitudes as described in his language versus as described in our language.)

Now, what is the relation between these three descriptive systems? The answer I will eventually defend is that (1) A and M should not really be considered two separate disciplines from a methodological perspective, ~~rather,~~ attitudes and meanings must be determined simultaneously, and (2) the joint A+M theory is an instrumental alternative to P. I will defend this claim by considering a number of alternative accounts which I do not think will work.

Can Meanings be Reduced to Attitudes?

One alternative that has been very popular is based on the hope that the methodological problem can be simplified by eliminating M by reducing it to A. By reducing meanings to propositional attitudes we are left with a single, hopefully more manageable problem, that of relating propositional attitude psychology to physical descriptions.

M Karl's meanings

Reduction

A Karl's Attitudes

Residual Methodological Problem

P Karl as a physical system

A significant number of important writers have argued that such a reduction is possible; the list includes Peter Strawson, H. P. Grice, Stephen Schiffer (in his earlier works), Jonathan Bennett and Brian Loar.[327]

Most advocates of this approach have taken their lead from Grice's classic 1957 article, "Meaning". That article contains a reductive program that is summarized in the following passage:

> (1) 'A meant-nn something by x' is (roughly) equivalent to 'A intended the utterance of x to produce some effect in an audience by means of the recognition of this intention'...
>
> (2) 'x meant something' is (roughly) equivalent to 'Somebody meant-nn something by x'...

---

[327] P. F. Stawson, "Meaning and Truth", in *Logico-Linguistic Papers*, (Methuen, 1971); H. P. Grice, "Meaning", *Philsophical Review*, 1957, 66, 377-388. Reprinted in J. F. Rosenberg and C. Travis, eds., *Readings ~~in the~~ Philosophy of Language*, (Prentice-Hall, 1971); S. Schiffer, *Meaning*, (Oxford University Press, 1972); J. Bennett, *Linguistic Behavior*, (Cambridge University Press, 1976); and B. Loar, *Mind and Meaning*, (Cambridge University Press, 1981).

> (3) 'x means-nn (timeless) that so-and-so' might as a first shot be equated with some statement or disjunction of statements about what 'people' (vague) intend (with qualification about 'recognition' to effect by x.[328]

(The English word 'meaning' is used in a variety of ways. Consequently, Grice uses 'meaning-nn' to stand for non-natural meaning as distinguished from "natural meaning", which is employed in the sentence 'Those clouds mean rain'.)

Here is how the Gricean strategy works. Assume that we have found out everything there is to know about the beliefs, desires, etc. of a community of people, but we don't know anything about the language that they speak. That is, we have a theory of propositional attitudes for the community, but we lack a theory of meaning. (Of course the relation between propositional attitude descriptions and physical descriptions must also be addressed by the Griceans. However, in this discussion I am focussing solely on the first step, that of reducing meaning to propositional attitudes.) The Gricean analyses given are intended to allow us to derive a theory of meaning from our knowledge of propositional attitudes. First, we identify cases where an individual is intending, by making some noise (or whatever), to produce in some audience some effect by means of recognition of this intention. This can be done entirely with the resources of our propositional attitude theory. Then by applying clause (1) above, we can characterize these cases as cases of a person non-naturally meaning something by his action. (This is now commonly called speaker-meaning.) Clause (2) allows us to say of the actions (utterances) themselves that they have meaning, and clause (3) says that if in the community a particular utterance is repeatedly used with a particular intention, then we can say that the utterance means such-and-such in a timeless sense. Advocates of this Gricean strategy have argued that David Lewis' work on linguistic conventions has helped make this last step a lot clearer (although Lewis himself does not subscribe to this reductionist doctrine).

---

[328] H. P. Grice, "Meaning", Philsophical Review, 1957, 66, 377-388. Reprinted in J. F. Rosenberg and C. Travis, eds., Readings in the Philosophy of Language, (Prentice-Hall, 1971), 442.

There is no doubt that this is a tempting strategy, and that there is a wealth of insight to be found in the works of the authors mentioned above. However, I think the strategy fails.

The question of whether or not this strategy is on the right track is one of the key issues in contemporary philosophy of language. Strawson writes:

> A struggle on what seems to be such a central issue in philosophy should have something of a Homeric quality; and a Homeric struggle calls for gods and heroes. I can at least, though tentatively, name some living captains and benevolent shades: on the one side, say, Grice, Austin and the later Wittgenstein; on the other, Chomsky, Frege, and the earlier Wittgenstein.[329]

Strawson sees the struggle primarily as one regarding the relation between concepts (does the concept of meaning presuppose the concept of communicative intention?) rather than one of methodology, but these are really two ways of approaching the same issue. In any case, Strawson's passage points to the centrality of this issue throughout the entire post-Fregean era.

---

[329] P. F. Strawson, "Meaning and Truth", in Logico-Linguistic Papers, (Methuen, 1971), 172.

The strategy of reducing meanings to attitudes fails, I think, because it cannot adequately address two problems which I shall call (a) the problem of intention identification, and (b) the problem of novel utterances. Each of these problems will be taken up in turn.

The Problem of Intention Identification

This problem casts doubt on clause (1) of Grice's reductive strategy. Donald Davidson states the problem this way:

> There can be no objection...to detailing the complicated and important relations between what a speaker's words mean and his non-linguistic intentions and beliefs. I have my doubts about the possibility of defining linguistic meaning in terms of non-linguistic intentions and beliefs, but those doubts, if not the sources of those doubts, are irrelevant to the present theme.
>
> The present theme is the nature of the evidence for the adequacy of a theory of interpretation. The evidence must be describable in non-semantic, non-linguistic terms if it is to respond to the question we have set; it must also be evidence we can imagine the virgin investigator having without his already being in possession of the theory it is supposed to be evidence for. This is where I spy trouble. There is a principled, and not merely a practical, obstacle to verifying the existence of detailed, general and abstract beliefs and intentions, while being unable to tell what a speaker's words mean. We sense well enough the absurdity in trying to learn without asking him whether someone believes there is a largest prime, or whether he intends, by making certain noises, to get someone to stop smoking by that person's recognition that the noises were made with that intention. The absurdity lies not in the fact that it would be very hard to find out these things without language, but in the fact that we have no good idea how to set about

> authenticating the existence of such attitudes when communication is not possible.[330]

Davidson is saying, then, that the Gricean strategy will work only if it is methodologically possible to identify complex Gricean intentions without bringing any semantic or linguistic considerations to bear. And this he maintains is impossible because communication is required in that very methodological exercise. That is, the speaker and the interpreter have to be able to communicate linguistically (at least at some level) if interpretation is possible. In claiming this Davidson is linking himself to the actual practises of anthropologists and field linguists who have always based their interpretative work on communication with informants; communication that is limited at first, and characterized by bold guesswork on the part of the interpreter, but which gets progressively more sophisticated as the enterprise proceeds. Davidson learned this method from Quine, who was in turn influenced by the work of the structuralist linguists.[331]

The last sentence in the passage quoted above states that it is not merely difficult in practise, but impossible in principle to identify propositional attitudes without communication. Actually Davidson is wrong on stressing communication. What is really essential for Davidson is that the identification of a person's propositional attitudes cannot be accomplished without knowing a lot about the person's attitudes towards sentences. But communication is only one way of finding out about a person's attitudes towards sentences. Imagine there is a Robinson Crusoe who is in the habit of always carrying a notebook. Robinson goes for daily walks, and he stops to enter sentences in his notebook whenever he finds something interesting. Suppose there is another person on the island, a mad anthropologist who has two goals: to remain unnoticed by Robinson, and to come up with a complete description of Robinson's propositional attitudes and his language. Each day the anthropologist follows Robinson around, noting the environmental conditions that prompt Robinson to make entries in his notebook. Each night the anthropologist sneaks into Robinson's hut and records Robinson's sentences, and then correlates them with the environmental conditions that prompted their writing. Now as I understand Davidson (including his writings beyond the passage quoted), this anthropologist would be able to interpret Robinson, provided that he had nightly access to his notebook, but without that access interpretation would be impossible. What is essential to Davidson is that the interpreter have access to the subject's attitudes towards sentences (more about this in section 4.3). Communication is one method of gaining access, but espionage will also work.

---

[330] D. Davidson, "Belief and the Basis of Meaning", Synthese, 1974, 27, 309-323. Reprinted in his Inquiries into Truth and Interpretation, (Oxford University Press, 1984), 143-144.

[331] This comment on Quine is somewhat speculative, but Quine's interest in the techniques of the structuralists seems reasonably apparent in his "The Problem of Meaning in Linguistics", in From a Logical Point of View, 2nd ed., (Harper and Row, 1961).

If Davidson is correct in his claim that belief identification requires knowledge of the subject's attitudes towards sentences then the Gricean reductive strategy will not work. Grice's goal is to analyze linguistic concepts in terms of the propositional attitudes, but Davidson is pointing out that propositional attitudes cannot be identified without assuming that the subject has attitudes towards sentences, which are linguistic entities.

Is Davidson's claim viable? It can be divided into two parts. (i) Davidson is claiming that it is possible to identify propositional attitudes through a method that begins with the evidence of a subject's attitudes towards sentences. (ii) Davidson is claiming that it is impossible to identify propositional attitudes without beginning with the subject's attitudes towards sentences. Now, I will be discussing (i) in more detail in section 4.3; what is important now is (ii), for it is (ii) that constitutes a refutation of the Gricean strategy, or any other strategy that falls under the umbrella I am calling "supervenience of propositional attitudes".

Jonathan Bennett, commenting on the passage by Davidson quoted above, points out that Davidson offers no real argument in favor of (ii), and considers Davidson's rejection of Grice a "breathtaking snub" [1976:271]. Bennett's book is a tour de force in working out the details of the Gricean strategy. Starting with a behavioral definition of belief, Bennett argues that we can assemble evidence that would demonstrate that a creature had met Grice's conditions for speaker-meaning, all without making use of linguistic concepts or assumptions. Bennett claims that he has shown, through hard work, how to do what Davidson glibly claims to be impossible.

It is true that Davidson fails to defend his negative point in the paper under discussion, but in one published the following year,[332] he is not so glib. In the latter paper Davidson argues that:

(i) A creature cannot have propositional attitudes unless he has beliefs, for belief is the central propositional attitude, constituting the background against which all other propositional attitudes are formed.

(ii) A creature cannot have beliefs unless he has the concept of belief. The concept of belief involves the contrast between truth and error.

(iii) The contrast between truth and error cannot emerge in a solipsistic setting. Rather, truth contrasts with error in the following way: truth is an objective, public standard, error is an individual deviation from the public standard.

(iv) But this contrast between public truth and individual error cannot be found in a community of languageless creatures. The contrast emerges only because of the activity of linguistic interpretation, where one creature is trying to understand the utterances of another.

It follows, then, that "a creature cannot have thoughts unless it is an interpreter of the speech of another".[333] As arguments go, this one is not for the squeamish. A reductionist like Bennett would no doubt be willing to go toe-to-toe with Davidson, especially on the key point, point (iv).

---

[332] D. Davidson, "Thought and Talk", in S. Guttenplan, Mind and Language, (Oxford University Press, 1975). Reprinted in Inquiries into Truth and Interpretation, (Oxford University Press, 1984).

[333] Ibid., 157.

My attitude toward Davidson's argument is one of caution. He may well be right, but I sense that these are weighty matters, and deserve careful attention. The argument is vaguely reminiscent of Wittgenstein's Private Language Argument, and the dust from that one is far from settled.

However, in the same paper Davidson considers two other supplementary reasons one might give for the thesis that thought requires language. One is that the paratactic analysis of belief, supplemented with Quine's theory of utterance mimicking, links the holding of a thought to a speech act. But Davidson points out that the paratactic analysis "is not an argument, but a proposal, and a proposal we need not accept."[334] Davidson's comments are appropriate, but even more damning is that the paratactic analysis need not imply that the thinker is a language user; it is only necessary that the interpreter be one. (The belief-error argument given above is much stronger; it says you cannot be a thinker unless you are an interpreter. The paratactic analysis implies, at most, that if you are an interpreter of beliefs you must be a language user.) This point is especially clear in Stitch's version of the paratactic analysis.

The second supplementary reason that Davidson considers is the following:

---

[334] Ibid., 167.

> ...[W]ithout speech we cannot make the fine distinctions between thoughts that are essential to the explanations that we can sometimes confidently supply. Our manner of attributing attitudes ensures that all the expressive power of language can be used to make such distinctions. One can believe that Scott is not the author of Waverly while not doubting that Scott is Scott; one can want to be the discoverer of a creature with a heart without wanting to be the discoverer of a creature with a kidney. One can intend to bite into the apple in the hand without intending to bite into the only apple with a worm in it; and so forth. The intensionality we make so much of in the attribution of thoughts is very hard to make much of when speech is not present. The dog, we say, knows that its master is home. But does it know that Mr Smith (who is his master), or that the president of the bank (who is that same master) is home? We have no real idea how to settle, or make sense of, these questions. It is much harder to say, when speech is not present, how to distinguish universal thoughts from conjunctions of thoughts, or how to attribute conditional thoughts, or thoughts with, so to speak, mixed quantification ('He hopes that everybody is loved by someone').
>
> These considerations will probably be less persuasive to dog lovers than to others, but in any case they do not constitute an argument. At best what we have shown, or claimed, is that unless there is behavior that can be interpreted as speech, the evidence will not be adequate to justify the fine distinctions we are used to making in the attribution of thoughts. If we persist in attributing desires, beliefs or other attitudes under these considerations, our attributions and consequent explanations of actions will be seriously underdetermined in that many alternative systems of attribution, many alternative explanations, will be equally justified by the available data.[335]

Although Davidson does not consider this a suitable argument to the conclusion that thought requires language, I personally am much happier with this than with the philosophically controversial truth-error argument. The basic point here is that without language we will have a crippled concept of belief. Davidson's remarks hint that if applied to languageless creatures our concept of belief will be crippled in two ways: we will not be able to understand the thinker as making fine discriminations in thought (e.g. believing that the demographic transition is correlated with industrialization versus believing that the demographic transition is caused by industrialization), and we will not be able to attribute intensional sensitivity to the speaker. I think it is very likely that if we insist on trying to attributing thought to languageless creatures we will have to recognize an even larger list of crippling considerations. For example, even that champion of languageless thought, Jonathan Bennett, despairs of attributing the capacity for logical inference to the languageless:

---

[335] Ibid., 163-164.

> We cannot have grounds for crediting a languageless creature with moderate logical acumen: either it can make no inferences, or it can make every inference; and in the latter case it believes everything which is entailed by any of its beliefs, which clearly seems to be absurd.[336]

It is easy to attribute thoughts to infants, animals and even computers; we do it all the time. However, the foregoing considerations show that unless the interpreted creature (machine) is also held to be a language user, the system of thought that we attribute to it will be a crippled version of our normal concept: it will be characterized by profound indeterminacy, we will not be able to draw the fine distinctions we normally do, we will not be able to make sense of the intensionality of thought, we will not be able to make sense of the creature reasoning about his thoughts, and the list may well be extended.

---

[336] J. Bennett, Linguistic Behavior, (Cambridge University Press, 1976), 116.

The Gricean has one move left. He can argue that we have to look at this as a bootstrapping operation. There is such a phenomenon as languageless thought, and it is badly "crippled" as outlined above. However, it is sufficient to support the Gricean mechanism, and therefore it is sufficient to get language going. Once language has been established in a community thought is enhanced.

This move is inspired partly by the age old appeal of all reductionist thought, but I think genetic considerations are a large part of the motivation. At one point in time there was no language on this planet. How did it come about? Surely this was a gradual process. The bootstrapping approach just outlined seems to provide an answer.

Here is an alternate story: Language and thought did not evolve. The underlying biological structures surely did, but language and thought, on the other hand, are timeless. At one point in the biological evolution of our species it became appropriate to speak of us as thinkers and language users, and before that it did not. But the appropriateness all comes in one fell swoop, not in incremental pieces. The ability to use a language and to think is part of a grand conceptual system for interpreting the behavior of certain creatures. But when we try to dismantle the system, and argue that it evolved bit by bit, the bits become almost incoherent, for the explanatory power of the system resides in its total overall coherence, not in self-sufficiency of individual pieces.

The claim I am making is that the attribution of full-blooded thought, that is, the attribution of a system of thought sophisticated enough to support the requirements of theory of action, requires the simultaneous attribution of competence over some language. For without language, thought is crippled in a variety of ways, and can no longer perform the theoretical role that we want it to perform ("the springs of action").

This claim is not a knock-down argument, but it seems reasonable enough to me, and certainly more reasonable than trying to construct an intentionally based social science on a crippled theory of thought. The claim has the somewhat shocking consequence that we will not be able to explain the historical evolution of language in a tribe using our intentional concepts. All we can say is that at one point in time the concepts clearly don't apply to the tribe, at a later point in time they clearly do apply, and that there is an intermediate period where we don't really know what to do. Similar remarks will apply to the acquisition of language in the child: we really can't have a theory of language acquisition (at least not one that answers all the possible questions) because if we tear language up into components, and propose non-linguistic patterns of thought that are precursors of language, then we end up with a bunch of elements that are crippled to the point where our grasp of our concepts is badly damaged.

I conclude, then, that we cannot identify thoughts in a languageless creature, at least not in a way that will meet the demands of a serious theory of intentional action.

The Problem of Novel Utterances

The previous problem cast doubt on clause (1) of Grice's reductive strategy. This one casts doubt on clause (3).

Paul Ziff has made this point as forcefully as anyone.

> (8) He's a son of a stickleback fish.
>
> (9) Sentence (8) means that the male referred to is a son of a small scaleless fish (family Gasterosteidae) having two or more free spines in front of the dorsal fin.
>
> ...I am inclined to think that (8) has been uttered only rarely. Nonetheless.... I am reasonably certain that (9) is a reasonably correct statement of the meaning of (8) and I am being reasonable in being so certain.
>
> If it is a correct statement of the meaning of (8)..., [then on Grice's account] (9) must be equivalent to some statement or disjunction of statements about what "people" intend to effect by (8). So Grice evidently maintains; for despite its vagueness, that is what is indicated by (iii) [i.e., clause (3) of Grice's analysis].
>
> But the question 'What do people intend to effect by (8)?' would not be a sensible question. Since hardly anyone has ever uttered (8) before, or so I suppose, one can hardly ask what people would intend to effect by it.
>
> ...[Grice's] suggestion is stated in terms of what people "intend", not in terms of what they "would intend". As such it obviously occasions difficulties with novel utterances. But matters are not at all improved by switching to what people would intend.
>
> For first, if a sentence is such that people in general simply would not utter it, then if they were to utter it, what they would intend to effect by uttering it might very well have nothing to do with the meaning of the sentence. What would a person intend to effect by uttering the sentence 'Snow is white and snow is white and snow is white and snow is white and snow is

> white'? I conjecture that a person uttering such a sentence would either be a philosopher or a linguist or an avant-garde novelist or a child at play or a Chinese torturer. What people would intend to effect by uttering such a sentence would likely have nothing whatever to do with the meaning of the sentence.
>
> Secondly, the switch to 'would' would be of help only if there were a constructive method of determining what people would intend to effect by uttering an utterance. There is no such method. There is not likely to be any (at least in our lifetime).[337]

In our discussion of Ziff's point, let us temporarily forget about the argument I previously developed: that no thoughts should be attributed to languageless creatures by serious intentional action theorists. For the purpose of this discussion, let us assume that it is possible attribute some thoughts, at least, without invoking linguistic concepts. So then it would be possible, through observation, to see cases of Gricean intentions (clause 1) in action, so we could construct a (finite) list of observed utterances and pair them with their Gricean meanings. But Ziff's point is that this is not yet an analysis of language, for language allows for novel utterances without limit. Even Bennett agrees that the term 'language' should not be applied to communication systems that have only a finite number of elements.

> I have shown how we could learn that some tribe use S to mean P, but only be observing how they use S in particular, and not, as one can in actual languages, by deriving the meaning of S from the meaning of other utterance-types though certain general principles. It is now time to take that further step. I think it is

---

[337] P. Ziff, "On H. P. Grice's Account of Meaning", Analysis, 1967, 28, 1-8. Reprinted in J. F. Rosenberg and C. Travis, eds., Readings in the Philosophy of Language, (Prentice-Hall, 1971), 448-449.

> the step from 'communication-system' to 'language', but I shall not insist upon that.[338]

(Of course I do not agree that Bennett has succeeded in the first step, that of identifying Gricean intentions independently of language, but as stated above, we will grant Bennett his point for the sake of discussion.)

What Ziff is questioning is how the Gricean can possibly take that step (clause (3) of the Gricean strategy). The Gricean has to introduce a constructive method for saying what the speaker would have meant if he were to have used a certain utterance, but the range of utterances considered must not be finite, but rather, is essentially unlimited. Now, one approach to this problem is obvious. We could modify clause (1) of the Gricean strategy to say that what is meant by the speaker is determined by the following list, and what follows is an infinite list of all the sentences of a language paired with their meanings. A list of T-sentences would do this. But since we cannot write down an infinite list, we will have to use some other technique to specify it. A Tarskian truth theory will do very nicely.

So we have solved the problem of novel utterances very nicely by referring to a Tarskian truth theory on the right-hand side of Grice's analysis of speaker meaning. But it doing so, we have completely given up on the Gricean reductive strategy, for we have denied that speaker-meaning can be defined independently of linguistic concepts, and therefore linguistic concepts cannot be defined in terms of speaker meaning.

---

[338] J. Bennett, Linguistic Behavior, (Cambridge University Press, 1976), 211.

In the passage quoted above, it is clear that Bennett thinks he can deal with novel utterances ('languages' as opposed to finite 'communication-systems') without compromising his Gricean scruples. Does he succeed? I do not think so.

Here is what Bennett does. First he constructs a "sentence-dictionary" based on "independent enquiries". An "independent enquiry", in Bennett's terminology, is the observation of a speaker making an utterance with Gricean intentions (that, according to Bennett, can be specified without invoking linguistic concepts). An independent enquiry will result in a fact of the form: S means P. A fact of this type becomes an entry in the sentence-dictionary. Obviously then, a sentence dictionary will be finite in size.

The sentence dictionary is all the evidence that Bennett needs. Everything from here on is based on analysis of the sentence dictionary. Bennett tries to show how the sentence dictionary provides evidence for the attribution of a first order semantic theory to an imaginary language he calls "Tribal".[339] However, Bennett does not even attempt to show how these semantic concepts and roles (i.e., that of name, predicate, demonstrative, quantifier, etc.) can themselves be constructed from or defined in terms of the sentence dictionary. Rather, Bennett assumes all this apparatus, and shows how, in his view, the sentence dictionary supports the thesis that this sentence component is a name for that item, that another component functions as a particular predicate, that a third is the existential quantifier, and so on. To repeat, the apparatus of semantic theory is assumed in Bennett's method; all the method does is guide the anthropologist in working out particular clauses of the theory for a particular language.

---

[339] *Ibid.*, chapter 8.

Given that this is how Bennett moves from a finite sentence dictionary to a theory that can account for novel utterances, can Bennett provide a satisfactory response to Ziff's issue? He cannot. What Ziff wants to know is how the Gricean would explain the meaning of a novel utterance. Bennett's answer has to be that the utterance means p, where p is the right hand side of a meaning giving consequence of a theory of meaning which is constructed out of primitive semantic concepts (e.g., name, predicate, demonstrative, quantifier, etc.) and which is supported by Gricean evidence for a finite subset of its consequences. In other words, Bennett has violated his own Gricean scruples for he has made use of linguistic concepts (name, etc.) to analyze the meaning of Ziff's novel utterance. He was not able to provide an analysis of the novel utterance's meaning using only psychological terms.

Bennett has failed, but might others succeed? The task would be to give a constructive procedure to produce Gricean intentions that did not turn out to rely on the productive power of a recursive semantic theory. My belief is that this is not possible (for I can't imagine how it could be done), but I have no proof that it is not. Instead I take the failure of Griceans to solve this problem after several decades of effort, as pretty good evidence that it is a bad strategy.[340]

The methodological approach we have been considering is that the theory of meaning can be reduced to the theory of propositional attitudes, which is, in turn, supervenient on the physical world. This approach has to be abandoned because the first step, the reduction of meaning to propositional attitudes, will not work. It will not work for two reasons: (a) any attempt to attribute thought to a languageless creature must necessarily result in a theoretically impoverished notion of thought, and (b) a theory of meaning must give the meanings of novel utterances, which means the Gricean must be able to give an account of counterfactual cases of speaker-meaning (i.e., what intention a speaker would have had if he had uttered this utterance) and it appears that this can only be done by introducing semantic concepts that have not been reduced.

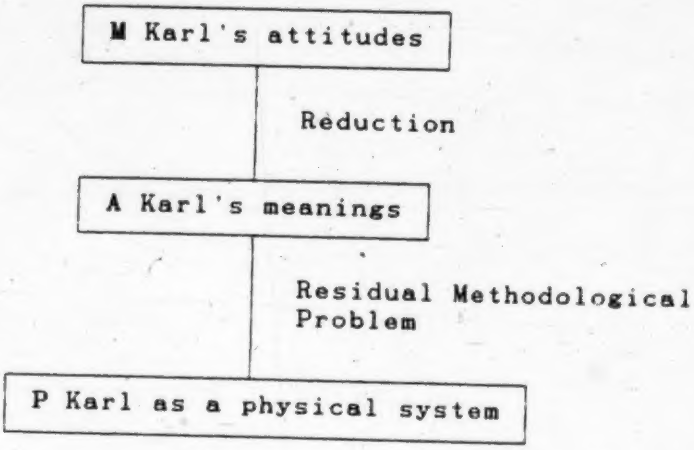
The conclusion we must draw, then, is that meanings cannot be reduced to attitudes.

Can Attitudes be Reduced to Meanings?

Another possibility is that attitudes can be reduced to meanings, which would reduce our overall problem to that of relating meanings to physical descriptions.

---

[340] Note that Stephen Schiffer, who was an important Gricean, has recently defected. See his Remnants of Meaning, (The M.I.T. Press, 1987).


M Karl's attitudes
Reduction
A Karl's meanings
Residual Methodological Problem
P Karl as a physical system

Has anyone even seriously proposed this? Perhaps the structuralist school can be interpreted as advocating such a reduction. Roland Barthes and Claude Lévi-Strauss have argued that linguistic theorizing is part of a larger study of symbols (sometimes called semiology) that includes mythology, kinship structures, rites and practises of various types, and other cultural phenomena.[341] In effect, these theorists are saying that these phenomena can be directly analyzed in terms of meaning (albeit a theory of meaning broader in scope than what we have been contemplating in these pages), rather than in terms of a propositional attitude theory (which is the more common approach, perhaps no better exemplified than in the work of Max Weber). At the level of individual psychology, Jacques Lacan has argued that Freudian psychology should be viewed as a theory of meaning (or perhaps he means it should be viewed from the vantage of a theory of meaning) rather than a theory about the mechanics of propositional attitudes (which is surely the way that it was intended by Freud).[342] Ray Birdwhistell has developed a theory of "kinesics" which analyses bodily movements on the model of structuralist linguistic theories (complete with "kines", "kinemes" and "kinemorphs").[343]

---

[341] Excerpts from Barthes, Lévi-Strauss, and others of their persuasion can be found in R. and F. DeGeorge, eds. The Structuralists from Marx to Lévi-Strauss, (Anchor Books, 1972).

I said that "perhaps" the structuralists can be interpreted as arguing that propositional attitude theory can be reduced to the theory of meaning. Alternatively, they may be saying that the theory of propositional attitudes should be eliminated; that the theory of meaning is all that social scientists need. Or they may deny that their doctrines have any relevance to the problem that I am addressing.

It is fair to conclude, I think, that the structuralists have not specified a clear program for either (a) the reduction of the propositional attitudes to the sphere of meaning, or (b) the elimination of propositional attitude explanations in favor of meaning explanations, although there seem to be tendencies toward one or the other in their work. I am not aware of any other attempts to reduce propositional attitudes to meaning, and furthermore, I cannot see how such an approach could possibly work. Consequently, I conclude (without further argument) that this alternative is not viable.

---

[342] J. Lacan, The Language of the Self, tr. by A. Wilden, (1968, reprint Dell Publishing, 1975).

[343] R. L. Birdwhistell, Kenesics and Context, (Ballantine Books, 1970).

Two Methodologies or One?

We have seen that we cannot simplify our method by reducing meanings to attitudes or vice versa. Consequently, both meanings and attitudes have to be related to the physical. But there are two ways of going about this. On the one hand, it might be the case that relating attitudes to the physical and relating meaning to the physical are two separate methodological problems. Or, it might be that there is a single method in which both attitudes and meanings are simultaneously related to the physical.

Figure 4.2.1
The Dual Method Alternative

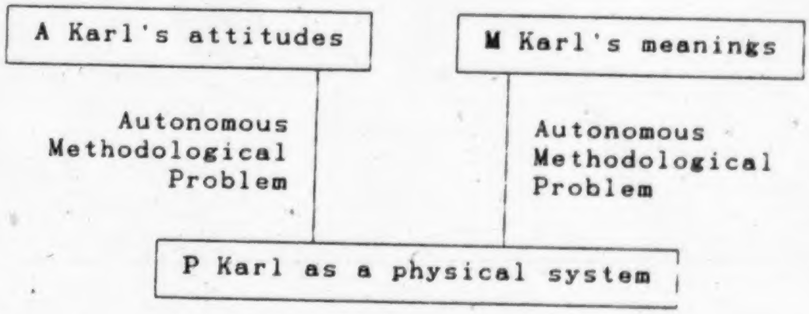
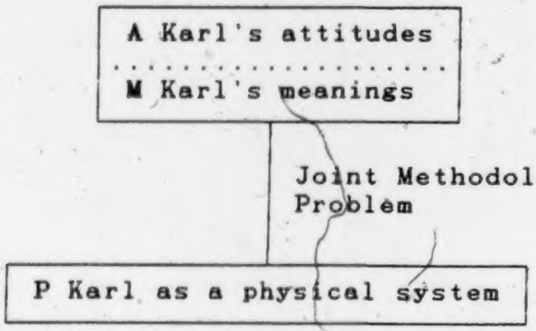

A Karl's attitudes
M Karl's meanings
Autonomous Methodological Problem
Autonomous Methodological Problem
P Karl as a physical system

Figure 4.2.2
The Single Method Alternative



My view is that the single method alternative is the correct one. I will demonstrate this by showing that neither of the two autonomous methodologies will work. But first some preliminary remarks. The arguments of Chapters Two and Three were that language and thought both have to be construed in the context of a single overall theory of action. If these arguments are valid, then the most obvious view of the methodological problem is this: In order to interpret a speech act, we must consider both aspects of our theory of action, viz., propositional attitudes and meaning. That means seeing the act as the production of an utterance in the context of certain intentions, but also seeing it as a species of action such that the act depends in some way on the content of the utterance. To see some behavior as a speech act means seeing it both in terms of its intentional structure and linguistic content (where "content" or "meaning" is exhausted by truth-conditions).

This suggests that the methodological problem is to simultaneously derive attitudes and meaning from the physical descriptions of individual pieces of verbal behavior. The burden of the argument is on the methodologist who sees it otherwise, i.e., who thinks that it is possible to solve for the two components independently. Note also that assuming the arguments I gave against reducing meaning to attitudes or vice versa are sound, then the methodologist who argues for independent methodologies has not made his case until he has shown how both independent methodologies would work. Let us consider the prospects are promising or not.

Imagine that we have travelled to some distant planet inhabited by intelligent creatures that communicate linguistically, although their language is not implemented through vocal-auditory systems, but through magnetic field generators and detectors. Our task is to come to know them as persons, which means coming to know their attitudes and meanings. Assume also that we believe that there are two autonomous methodologies, one that will allow us to determine attitudes from the physical evidence, and one that will allow us to determine meanings from the physical evidence. Consequently, we decide to break into two teams: the A-team and the M-team. The A-team is to return with a description of the attitudes of a representative group of the planet's inhabitants. The M-team is to return with a description of the meanings (i.e., truth conditions) of the sentences that the inhabitants have been observed to utter. Both teams have been ordered not to waste any time solving the other team's problem.

Will the teams succeed?

The Failure of the A-Team

The A-team's best strategy is to begin with assumption that the creatures are the result of the same sort of evolutionary process that characterizes life on earth. Assuming that the creatures do not appear to be undergoing a major environmental crisis we can make the further working assumption that they are reasonably well adapted to their environment. Consequently their patterns of avoidance of certain environmental states, and the pursuit of others, can be taken as evidence of preferences, i.e., of desires. With knowledge of these fundamental desires (they will no doubt have to do with food, shelter, self-preservation, reproduction, and so on), the A-team could no doubt make some further progress by identifying beliefs that reveal themselves as the creatures pursue their goals. For example if the creatures are seen to methodically construct shelters as the cold season approaches, it will no doubt be plausible to attribute many beliefs to them.

But these creatures are intelligent language-users. They not only have the skills to build shelters, they also have a richly developed mathematics, and their culture includes a strong oral tradition, and many legends are shared between them, and they are quick to offer moral advice to one another.

In order to complete their job, the A-team must account for the mathematical beliefs of the creatures, as well as their beliefs about the fictional characters and events of their legends, as well as their moral beliefs. How will the A-team fare with these more abstract beliefs?

Not well at all, I think it is safe to say. Davidson asks how we would know that someone believes that Napoleon is a great general other than by asking him (or in some way making use of his language to help identify his belief). Even a Gricean like Bennett would agree that the attribution of the bulk of the creatures' abstract beliefs will not be possible until after their language has been identified.

The Failure of the M-Team

The M-team will have even less success. Given the restriction that the M-team is not to make any assumptions about the creatures" propositional attitudes, the team must view the creatures as some sort of machines that emit magnetic fields as a function of their environmental conditions. But let us assume that the creatures, like us, not only speak of what they believe to be true about the environment, they sometimes also tell lies. Assume also that some of what they say is intended as fiction. Whether a magnetic field sentence is "uttered" sincerely, as a lie, or as a fiction will, of course, make all the difference to how it should be interpreted. The M-team is specifically restricted from making any distinctions between truthful assertions, lies and stories, since these can only distinguished by attributing intentions of various sorts to the creatures.

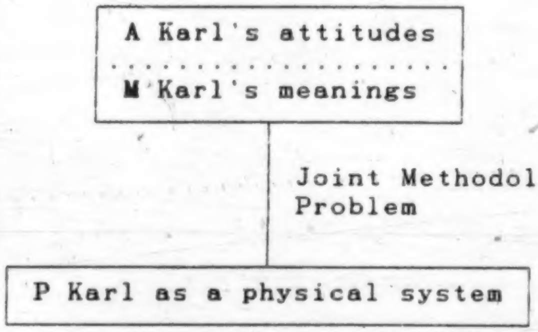
Indeed, without attributing propositional attitudes we cannot even make sense of the idea of attributing assent or dissent (to a sentence) to the creatures, and therefore yes-no questioning, a time-honored technique of anthropologists and field linguists, will not be available to the M-team.

The conclusion that we must draw is that the strategy of dividing up into two teams is a bad one. The attribution of attitudes and meanings should be conducted as a single methodological exercise in which thought and language are simultaneously identified.

Supervenience or Instrumental Alternatives?

The previous arguments have attempted to show that there is a single methodological problem, that of jointly determining attitudes and meanings from a basis of physical evidence.

Figure 4.2.3
The Single Method Alterative (Again)


A Karl's attitudes
M Karl's meanings
Joint Methodological Problem
P Karl as a physical system

We know from arguments given in Chapter Three that the relation between physics and attitudes/meaning cannot be one of reduction. So is the relation one of supervenience or should we view the physical perspective and the attitude/meaning perspective as instrumental alternatives?

The answer, I believe, hinges on how we view intentional explanations of actions: on whether we follow the "causal-realist" view of someone like Stephen Stitch, or whether we follow the "rational-instrumental" view of someone like Daniel Dennett.

An example of an intentional explanation of an action is the following: Donald flipped the switch because he wanted the light on and believed that by flipping the switch the light would go on. A "causal-realist" believes that this explanation works because the intentional state of Donald's wanting/believing is token-token identical with some physiological state that caused his finger to flip the switch. A "rational-instrumentalist" believes that the explanation works because all the components, i.e., Donald's wanting, his believing, and his subsequent action, hang together in a manner that conforms to some sort of norm of rationality. The "rational-instrumentalist" accepts that some wantings and believings may be token-token identical with physiological happenings, but is not especially troubled if they are not. In fact, the "rational-instrumentalist" says that it is possible to generate an unlimited number of consequences of any intentional explanation (e.g., Donald must believe that there is a bulb in the socket, that the electrical circuit is not damaged, etc.) in order for the norm of rationality to be satisfied, and sooner or later it will become implausible to insist that each one of these supporting beliefs and desires has a physiological counterpart.

As explained in the last chapter, I lean toward the instrumentalist view rather than the causal-realist view. Consequently, I hold that intentional descriptions and physical explanations are instrumental alternatives.

4.3 RADICAL INTERPRETATION

"Radical interpretation" is a term coined by Donald Davidson to refer to the task of jointly determining the attitudes and meanings of the members of a speech community from a physical description of their activities in their environments. The term derives from Quine's "radical translation" which is a similar although somewhat more restricted undertaking.

In a broad sense, the purpose of radical interpretation and radical translation is somewhat similar to the purpose of the discovery procedures that were developed by linguists such as Bloomfield and Harris. The similarity is the emphasis on starting only with uninterpreted physical evidence, and proceeding toward linguistic and psychological description only when it is explicitly warranted by the method. In the practical sense, however, the literature of radical interpretation is very different from the discovery procedure literature. The latter is intended as a set of practical procedures, hopefully to be used by field linguists on their next expedition. The methods and procedures of the radical interpretation literature are not intended as guidelines for actual practise, however. Writers contributing to the radical interpretation literature will assume that there is more physical evidence available than anyone is likely to gather. They will postulate methods that involve more successive iterations of some sub-method than anyone is likely to accomplish, and so on. The reason that the radical interpretation literature is so impractical is that the authors contributing to it are not directly interested in the practical problem facing anthropologists, rather, they are interested in the relation between physical descriptions and the language of attitudes and meanings, and they are exploring this question through the fiction of researcher who is not hampered by any practical concerns. This focus is nicely illustrated in the following passage from David Lewis.

> It should be obvious by now that my problem of radical interpretation is not any real-life task of finding our about Karl's beliefs, desires and meanings. I am not really asking how we could determine these facts. Rather: how do the facts determine these facts? By what constraints, and to what extent, does the totality of physical facts about Karl determine what he believes, desires and means? To speak of a mighty knower, who uses his knowledge of these constraints to advance from omniscience about the physical facts P to omniscience about the other facts determined thereby, is a way of dramatizing our problem - safe enough, so long as we can take it our leave it alone. The real-life knower has all the problems of our fictitious knower, and more besides: he does not have all of P to

> draw on, and he may be limited in endurance, intelligence or memory.[344]

Our concern is with the Linguistic Relativity Hypothesis. Our concern is whether or not the Linguistic Relativity Hypothesis is coherent, and if so, whether it is true. Now the Linguistic Relativity Hypothesis might be true, but it could be that it would be beyond the levels of endurance, intelligence or memory of any researcher to actually demonstrate it to be so for any particular pair of speech-communities. In other words, the truth of the Linguistic Relativity Hypothesis does not hinge on the practical limitations of any particular researcher. However, if an idealized researcher, unlimited in endurance, intelligence or memory, could not show the Linguistic Relativity Hypothesis to be true, then it could not be true. (A more careful way of putting this would be that the Linguistic Relativity Hypothesis would not be an acceptable scientific statement under these circumstances.) What this means is that the truth or coherence of the Linguistic Relativity Hypothesis is closely tied to the problem of radical interpretation as defined by authors like Donald Davidson and David Lewis. If the method of radical interpretation warrants the attribution of incommensurable attitudes to different groups of people, the Linguistic Relativity Hypothesis has a chance. If the method requires that all people share a common world view, then the Linguistic Relativity Hypothesis must be abandoned.

---

[344] D. Lewis, "Radical Interpretation", Synthése, 1974, 23, 331-344. Reprinted in D. Lewis, Philosophical Papers: Volume I, (Oxford University Press, 1983), 110-111.

Davidson on Radical Interpretation

Donald Davidson has proposed a method of radical interpretation that can be characterized as follows:

1. The method assumes that semantic and mental phenomena are supervenient on physical phenomena, but not (type)reducible to physical phenomena.

2. The method is consistent with the claim that "meaning ain't in the head"; that is, the method is consistent with the idea that a mental or semantic fact may be supervenient on a piece of the physical world that is not bounded by the skull of any particular individual.

3. Since reduction is rejected by Davidson, his method must allow for the recognition of intensional (i.e., semantic or mental) facts as a primitive capability of the researcher. I mean "primitive" in the sense that the researcher is able to recognize these facts immediately, that is, the researcher does not infer that the intensional facts are as they are on the basis of the recognition of certain non-intensional facts, but rather, recognizes the intensional facts directly without any inference from non-intensional facts.

4. The admission of the direct recognition of intensional facts may sound like Davidson is offering no method at all, for it may sound like he is saying that all you have to do is put on your intensionality-tinted sunglasses and go out and observe. But Davidson's position is more complex, and methodologically sound, than that. Davidson claims that there is a subset of intensional attributions that (1) is relatively non-controversial in terms of when and how it should be empirically applied, and

(2) from which the full set of intensional facts about a person can be derived.

5. Finally, Davidson does not recommend his method as a practical guide for social scientists, but rather, as an explication of the relation of the intensional idioms to the physical world.

Davidson's key methodological principle is that as social scientists, we must have a primitive capability to recognize intensional facts (since reduction is not possible) but that there is a subset of such facts that are relatively clearer to recognize than the others, and from which the others can be inferred. This is an approach that Davidson learned from Frank Ramsey's work on the measurement of subjective probability; an intellectual debt that Davidson repeatedly points out.[345] It will facilitate the discussion to briefly review Ramsey's work.

Ramsey was attempting to explicate the notion of probability, but as he was working in the subjectivist school, his work can be viewed as a contribution to the methodology of decision theory, which in turn can be viewed as a formalization of the folk theory of rational action. Decision theory predicts action as a function of subjective probability (weighted belief) and a utility function (weighted desire). To apply decision theory to a subject we need first to know his weighted beliefs and desires. But this is not so easy, for all that we can directly observe are actions. Belief and desire are the invisible hypothesized factors behind the action. So in order to identify beliefs and desires it would seem that we have to observe action, and abstract the belief and desires out of it. But the problem, as Ramsey and a legion of early economists recognized, is that the theory views any action as the resultant of belief and desire and accordingly different mixes of belief and desire can result in the same action. (For example, Jack's apple-eating can be explained either by "Jack ate Jill's apple because he was hungry and believed that eating it would alleviate his condition" or "Jack ate Jill's apple because he wanted to cause Jill grief and he believed that eating her apple would cause her grief". Similarly, Jane's lottery-playing can be explained either by "Jane plays the lottery because she believes [falsely] that the odds of her number being drawn are such that she will probably come out ahead in the long run" or "Jane plays the lottery because she has a non-linear subjective utility function for money". And so on.)

---

[345] See D. Davidson, "Belief and the Basis of Meaning", *Synthése*, 1974, *27*, 309-323. Reprinted in D. Davidson, *Inquiries into Truth and Interpretation*, (Oxford University Press, 1984); and D. Davidson, "Toward a Unified Theory of Meaning and Action", *Grazer Philosophische Studien*, 1980, *11*, 1-12.

Ramsey showed a way to overcome this indeterminacy and to assign a unique set of weighted beliefs and desires to an individual. His method is based on the study of gambling preferences which can be used to solve for belief. Once belief is established then it is possible to solve for desire by studying action (with belief held constant, desires can be read off the actions, assuming the agent is rational).

Suppose there is a gamble as to whether p or not-p is true, and the payoff is the same for either possibility. If a subject is indifferent between which side of the gamble he prefers (p or not-p) then the subjective probability of p is 0.5 for that subject. More generally, by studying the total pattern of preferences among gambles (and concentrating on the cases toward which the subject is indifferent) the researcher can assign subjective probabilities to the entire range of beliefs under consideration.

Once subjective probabilities have been established, it is a simple matter to determine the subject's utility function through the study of his actions, since assuming that the subject is rational, desires are a function of actions and beliefs.

Ramsey's method does not reduce belief and desire to the physical world. Rather, it reduces belief and desire to a single, more methodologically tractable set of propositional attitudes: ordinal preferences between gambles.

Here is what Davidson has to say about this:

1. On the positive side, Ramsey's general method is the correct one for the empirical identification of intensional phenomena. That is, since reduction is not possible, what we must do is reduce complex systems of intensional phenomena to simpler subsets that are more methodologically tractable.

2. On the negative side:

(a) Davidson holds that Ramsey's method is incomplete as it stands, for we cannot suppose that we could find out much about a subject's gambling preferences unless we could talk to him. That means we have to be able to interpret his utterances; in other words we need a theory of meaning for his language. We need a broader method, then, one that addresses meaning as well as belief and desire.

(b) Davidson points out that Ramsey's method is ontologically obscure, as it considers propositions as the objects to which subjects assign subjective probability. (A belief, on Ramsey's view, is a three-place relation between a person, a proposition and a number.)

In contrast to Ramsey, Davidson sees the methodological problem more broadly. According to Davidson, we must simultaneously solve for belief, desire and meaning. Also contrasting with Ramsey in terms of his ontological preferences, Davidson suggests that we deal with sentences rather than propositions wherever the need arises for such objects. However, Davidson agrees with Ramsey that we have to identify some methodologically tractable subset of what we are after, and that we have to show how what we are after can be derived from that subset.

So what is Davidson's methodologically tractable subset and what is his derivation procedure? He gives two answers, an early approach largely indebted to Quine, and a later approach that owes a heavy debt to Richard Jeffrey. The early approach is discussed in a number of Davidson's papers. The later approach is hinted at in one paper, and briefly outlined in another. In the introduction to his 1984 volume of essays he states that he is working on a more detailed exposition of his methodology, one that presumably would develop the later approach, but so far this work has not appeared.

Both Davidson's early and later approaches will now be outlined.

Davidson's Early Method

The methodological foothold that Davidson proposes is that of a subject's holding a sentence true under certain causal conditions. Davidson presumes that the investigator could determine that a person holds a particular sentence true without knowing what the sentence means. By accumulating enough facts of the form:

S holds s true under circumstances c

Davidson claims that we can determine both meaning (i.e., a Tarskian truth theory for the subject's language) and the subject's beliefs. Once belief has been determined, the subject's desires can be transparently read off his actions (under the assumption that he is acting rationally, of course).

So assuming we have assembled the facts that are available to us, i.e., facts of the form:

S holds s true under conditions c

how do we go from there to a Tarskian truth theory and a set of belief attributions? In his paper "Radical Interpretation" (which is his basic statement of the early approach to radical interpretation) Davidson outlines the following steps:

1. We collect facts of the form

(E) Kurt belongs to the German speech community and Kurt holds true 'Es regnet' on Saturday at noon and it is raining near Kurt on Saturday at noon.

2. We examine the pattern of facts of the form (E) and determine generalizations of the form:

(GE) (x)(t)(if x belongs to the German speech community then (x holds true 'Es regnet' at t if and only if it is raining near x at t))

3. Then we apply the principle of charity to obtain relativized T-sentences like

(T) 'Es regnet' is true-in-German when spoken by x at time t if and only if it is raining near x at t.

4. We determine the logical form of all sentences by studying "the class of sentences always held true or always held false by almost everyone almost all of the time (potential logical truths) and patterns of inference."[346] The point of this step is to identify "predicates, singular terms, quantifiers, connectives and identity". This passage reflects Davidson's bias toward viewing natural languages as being first order languages, but in any case, the purpose of the fourth step is to determine matters of logical form.

5. We construct a referential structure by examining how the truth verdicts of indexical sentences vary according to speakers/times and associated circumstances.

6. We work the analysis through for all the remaining sentences in the language.

Davidson's Later Method

In a paper written a year after "Radical Interpretation" Davidson devotes one paragraph to an alternate approach.[347] The key idea of this earlier approach is that we begin by studying what our subjects hold-true. The alternate approach is to study what sentences our subjects prefer-true. Davidson barely hints at why he wants to change his basic notion. It seems to me that he thinks that it would be easier for the researcher to observe instances of preferring-true compared to observing instances of holding-true. He also alludes to Richard Jeffrey's theory of belief/desire attribution. Jeffrey's theory is a competitor to (and in Davidson's opinion an improvement on) Ramsey's theory. Jeffrey's theory builds up measures of subjective probability and value from the initial data of what propositions a subject prefers-true (Ramsey's used preferences between gambles as the fundamental data). One reason why Davidson prefers Jeffrey's theory is that it is less "ontologically murky". So it is not entirely clear why Davidson stated a preference for Jeffrey's theory in 1974: it may have been for methodological reasons, it may have been for ontological/theoretical simplicity, or it may have been for both reasons.

---

[346] D. Davidson, "Radical Interpretation", Dialectica, 1973, 27, 313-328. Reprinted in D. Davidson, Inquiries into Truth and Interpretation, (Oxford University Press, 1984), 136.

[347] D. Davidson, "Belief and the Basis of Meaning", Synthése, 1974, 27, 309-323. Reprinted in D. Davidson, Inquiries into Truth and Interpretation, (Oxford University Press, 1984), 148.

In any case, the idea was developed further in a paper published a few years later.[348] In that paper Davidson gives a very straight forward reason why one might be unhappy with the holding-true approach. The problem with holding-true is that it is an all-or-none thing. This causes us a problem in determining the logical form of the sentences of a natural language. For as stated in point 4 of the earlier method, one of our key sources of evidence for logical form is the evidential relations between sentences. We can make limited progress in this by studying changes, over time, in the patterns of sentences held-true by an individual. That is, we would study how changes in the truth verdict that an individual makes regarding a particular sentence correlates with changes in his truth verdicts regarding other sentences. But this sort of evidence will only get us so far; it would be much better, according to Davidson, if we could study how patterns of degrees of belief change. And here is where taking preferring-true as our fundamental notion can help, for then we can exploit Jeffrey's theory of how to build up weighted beliefs from the study of preferring-true.

---

[348] D. Davidson, "Toward a Unified Theory of Meaning and Action", Grazer Philosophische Studien, 1980, 11, 1-12.

I will not present Davidson's later method in detail, for the following three reasons:

1. The method is still under development by Davidson. He has promised a book on the subject, and what appears in the 1980 article is merely a sketch.

2. Davidson has written a number of articles about the relevance of the radical interpretation problem to various issues in philosophy and social science. One of these articles is directly relevant to the Linguistic Relativity Hypothesis.[349] However, all of these articles are based on the earlier method of holding-true, rather than the later method of preferring-true.

---

[349] D. Davidson, "On the Very Idea of a Conceptual Scheme", Proceedings and Addresses of the American Philosophical Association, 1974, 47. Reprinted in D. Davidson, Inquiries into Truth and Interpretation, (Oxford University Press, 1984). This paper is discussed in detail in section 4.4.

3. Finally, and most importantly, the actual step-by-step methods that are proposed for the solution of the radical interpretation problem are probably not that relevant for issues like the Linguistic Relativity Hypothesis in any case. What is important are the constraints that govern the radical interpretation problem. At least that is what David Lewis argues, and I agree with him. Our review of Davidson's step-by-step approach has hopefully developed an appreciation of the issues in the radical interpretation problem. Let us now examine how Lewis looks at the issue.

Lewis on Radical Interpretation

In the two articles he devoted to the subject, Donald Davidson approached the issue of radical interpretation as follows: he tried to outline the steps that an ideal interpreter (unlimited in endurance, etc.) would have to take in deriving attitudes and meanings from the physical evidence. David Lewis, however, points out that Davidson's focus on outlining a series of steps is not appropriate, given that the purpose of clarifying the problem of radical interpretation is not to solve any practical problem. Since the purpose of clarifying the radical interpretation problem is simply to shed light on the extent to which physical facts determine the psychological and semantic facts, Lewis says that all we really need to focus on are the constraints that must be observed in the construction of any step-by-step method. If we understand the constraints that any method has to observe, then we understand everything there is to know about how the physical facts determine the other facts. (Indeed, an omniscient knower with plenty of time on his hands need not follow any method, he just has to rank all the possible psychological/semantic descriptions of Karl in terms of how well they conform to the constraints between psychological/semantic descriptions and physical descriptions. Step-by-step methodologies are for mere mortals.) So, for certain types of issues it is constraints that matter, not methods. And for the Linguistic Relativity Hypothesis it is the constraints that matter, for what we are concerned about is whether people can have radically different beliefs and languages, not whether researchers with limited endurance, etc. can detect this.

Lewis has presented a list of constraints that he believes apply to the problem of radical interpretation (although he states with candor that he is not sure if the list is complete or if it is redundant). His constraints are the following six:

1. The Principle of Charity
2. The Rationalization Principle
3. The Principle of Truthfulness
4. The Principle of Generativity
5. The Manifestation Principle
6. The Triangle Principle

It will be helpful, in explaining Lewis' six constraints, to review his diagram of the radical interpretation problem:

P Karl as a physical
system

Ao Karl's attitudes:
beliefs and desires
as expressed in
our language

Ak Karl's attitudes:
beliefs and desires
as expressed in
Karl's language

M Karl's meanings:
truth conditions of
his full sentences
....
(and denotations, etc.
of constituents of
sentences)

[350]

The six constraints can be characterized, using Lewis' diagram, as follows:

1. **The Principle of Charity**

This principle "constrains the relation between Ao and P: Karl should be represented as believing what he ought to believe, and desiring what he ought to desire."[351] Lewis points out that this principle can be interpreted in two ways. (a) We can take it to mean that treating Karl charitably requires attributing to him mostly true beliefs and appropriate desires, where our beliefs and desires are considered the yardstick of what is true and appropriate. (b) We can take it to mean that both Karl and we form our beliefs and desires according to the same principles, and that therefore what Karl believes and desires would be what we would believe and desire if we were in his shoes. In what follows I will assume that the principle of charity should be interpreted in sense (a). Interpretation (b) states an interesting point: that Karl cannot form his beliefs and desires in any old way; there are constraints on the way he forms his beliefs and desires. I think that we should recognize this point, but we should subsume it under the idea that Karl should be viewed as a (largely) rational actor, and that he forms his beliefs and desires (largely) rationally. Thus, I think the point that is being made in (b) should be moved to Lewis' principle number 2 (below), where he talks about rationality.

---

[350] D. Lewis, "Radical Interpretation", *Synthése*, 1974, 23, 331-344. Reprinted in D. Lewis, *Philosophical Papers: Volume 1*, (Oxford University Press, 1983), 108.

[351] *Ibid.*, 111.

But why should we endorse the principle of charity, interpreted in sense (a), at all? Donald Davidson argues that charity, in this sense, is a necessary prerequisite to interpretation. This argument will be presented in section 4.4.

2. The Rationalization Principle

This principle "constrains the relation between Ao and P: Karl should be represented as a rational agent; the beliefs and desires ascribed to him by Ao should be such as to provide good reasons for his behavior, as give in physical terms by P."[352] This principle obviously requires more explanation of the idea of rational agency, and of the idea that beliefs and desires can provide a "reason" for behavior. Lewis answers these questions by alluding to decision theory, which in his view is just an elaboration of the notions of rationality and reason-giving that we find in folk psychology. (Davidson and Lewis are of a like mind on this.)

---

[352] Ibid., 113.

The principle of rationality as stated by Lewis is a little too narrow. Part of our notion of rationality is that when we form or change our beliefs or desires we must do so in a rational manner. So besides enforcing a relation between attitudes and behavior (as stated above by Lewis), the principle of rationality should also enforce an internal relation on attitudes. Lewis recognized this issue when discussing his interpretation (b) of the principle of charity, but the issue is more properly placed as a part of the rationality constraint. Again, it is possible to appeal to decision theory for more elaborate models of belief and desire formation.

3. The Principle of Truthfulness

This principle is perhaps unique to David Lewis. Lewis has developed a theory of linguistic conventions in which the relation between a speech-community and the language that they use is characterized as a convention of truthfulness and trust in that language. This theory involves the attribution of a great many beliefs and desires to each of the members of the speech-community, and many of these beliefs and desires are about linguistic utterances (their appropriateness in various situations, what other people think about them, etc.). Consequently, Lewis believes that his theory of linguistic conventions constrains the relation between **M** and Ao. Specifically, once **M** is determined, the theory of linguistic conventions will allows us to derive a number of details about Ao.

4. **The Principle of Generativity**

This principle "constrains M: M should assign truth conditions to the sentences of Karl's language in a way that is a least finitely specifiable, and preferably also reasonably uniform and simple."[353] This principle has been endorsed by many philosophers of language, and of course has been central to linguistics since Chomsky's Syntactic Structures.

5. **The Manifestation Principle**

This principle "constrains the relation between P and Ak, and to a slight extent also Ao. Karl's beliefs, as expressed in his own language, should normally be manifest in his dispositions to speech behavior. The sentences (in context) that he could be made to utter should normally be among those that express propositions that he believes to a high degree."[354] This principle is endorsed by most speech act theorists. I believe the principle to be valid, but Lewis' phrase "should normally be manifest" is somewhat ambiguous. "Normally" might mean that the majority of Karl's utterances express his beliefs, or it may mean that all his utterances in "normal" speech acts of assertion express his beliefs, although this type of utterance may not constitute the majority of his utterances in a statistical sense. I think only the latter is defensible.

---

[353] Ibid., 115.

[354] Ibid., 115.

In any case, the manifestation principle should be viewed as a consequence of whatever flavor of speech act theory one endorses. Indeed, there are other consequences of speech act theory that can be brought to bear as constraints on the problem of radical interpretation. If one holds that the Gricean model of speaker's intentions is more-or-less correct, then each speech act will allow the interpreter to read off number of beliefs and desires besides a belief in the proposition of the sentence expressed.

6. The Triangle Principle

This principle "constrains the three-way relation between Ao, M and Ak: Karl's beliefs and desires should be the same whether expressed in his language or ours."[355]

I will now make some special comments on the second principle, the principle of rationalization.

In the closing paragraphs of section 4.2 I briefly mentioned a difference of opinion that exists between writers like Stephen Stitch and Daniel Dennett on the topic of intentional explanations of actions. Stitch believes that explanations like 'Donald flipped the switch because he wanted the light on' work because they identify a wanting (which is identical to some physiological state) that caused the flipping of the switch. Dennett believes that the explanation works because it conforms to a normative pattern characteristic of intentional explanation, even though there may be no one to one correlation of intentional states with physiological states. Now, I claim that Lewis' second constraint, the principle of rationality, puts him in the Dennett camp rather than the Stitch camp. (I say this even though Lewis explicitly endorses Stitch's causal-realist view in some of his other arguments. Lewis, like Davidson and many other authors, vacillates between the two approaches to intentional explanation. Stitch and Dennett are to be commended for sharpening the distinction and showing the extent to which the two views are inconsistent.)

---

[355] Ibid., 115.

The principle of rationality states that Karl's beliefs and desires should be such as to explain his behavior by providing reasons for it, and his beliefs and desires should be rationally formed. By appealing to decision theory as a further elaboration of these notions, Lewis is in effect endorsing Dennett's position, since decision theory is a normatively based calculus that can be imposed globally as a model of someone's behavior, but it is extremely unlikely that every belief, desire, calculation and consideration that is postulated in decision theory will actually be mirrored in the subject's brain in some way. Decision theory provides an instrumental alternative to the physiological explanation of the subject's activities, but it cannot be tightly correlated with a physiological explanation at a fine level.

Since Stitch disagrees with this approach to intentional explanation he would therefore be unhappy with Lewis' principle of rationality. Stitch would agree that we require a principle that connects Karl's beliefs/desires with his actions and his beliefs/desires with subsequently formed beliefs/desires, but he would formulate it something like this:

2'. The Causal Principle

This principle constrains the relation between Ao and P: Karl's beliefs and desires should be represented as having a causal role in his behavior, and as having a causal role in the formation of subsequent beliefs and desires.

Stitch would not be happy just to add this principle to Lewis' list. Rather, on his view it should replace Lewis' second principle, the rationalization principle. To summarize: theorists like Lewis or Dennett who see psychological explanations as instrumental alternatives to physical explanations (held together with a normatively based rational-actor model) will endorse the principle of rationality as a constraint on radical interpretation. On the other hand, theorists like Stitch who see psychological explanations as invoking causal relations between mental states and behavioral states (and therefore requiring a token-token identity of mental states with physical states) will reject the rationality principle and replace it with a causal principle. (It is worth repeating that many theorists, Lewis and Davidson among them, simultaneously endorse both constraints. I believe, with Stitch and Dennett, that this is inconsistent.)

For reasons given in Chapter Three, I endorse the rationality principle and not the causality principle. Consequently, the causality principle will be presumed to play no role in radical interpretation.

4.4 DAVIDSON ON CONCEPTUAL RELATIVITY

Donald Davidson's paper "On the Very Idea of a Conceptual Scheme" is a very important contribution to the issue of conceptual relativity in general, and the Linguistic Relativity Hypothesis in particular.[356] Davidson is against the idea of conceptual relativity. He is out to demonstrate two things: (1) that the idea of incommensurable conceptual schemes makes no sense, and (2) consequently, the very idea of a conceptual scheme makes no sense.

Davidson's argument obviously has relevance to the Linguistic Relativity Hypothesis, but he also is directing it at the doctrines of Kuhn, Feyerabend and other historians and philosophers of science who argue that scientific history is characterized by conceptual revolutions such that the post-revolutionary conceptual scheme is incommensurable with the pre-revolutionary scheme.

Davidson asks how we might make sense of the idea of conceptual schemes being incommensurable. He suggests that we identify schemes with languages and then equate the notion of incommensurability with non-translatability.

---

[356] D. Davidson, "On the Very Idea of a Conceptual Scheme", Proceedings and Addresses of the American Philosophical Association, 1974, 47. Reprinted in D. Davidson, Inquiries into Truth and Interpretation, (Oxford University Press, 1984).

> We may accept the doctrine that associates having a language with having a conceptual scheme. The relation may be supposed to be this: where conceptual schemes differ, so do languages. But speakers of different languages may share a conceptual scheme provided that there is a way of translating on language into the other. Studying criteria of translation is therefore a way of focussing on criteria of identity for conceptual schemes.[357]

In other words, if there is a pair of languages that cannot be translated into one another, then we can say of the respective speech-communities that they have incommensurable conceptual schemes. Davidson's proposal has the virtue that it allows us to focus on translation, an issue that seems a lot more tangible than the clouded notion of "incommensurability". However, is Davidson correct in identifying schemes with languages, or has he merely changed the subject? As first glance it might appear that it is the latter, especially if you consider the incommensurability thesis of Kuhn who argues that there can be incommensurability between scientific theories, even though both theories are expressed in a single language, say English. However, it would be wrong to invoke this line of thinking against Davidson, for it would be based on a mis-reading of Kuhn. Kuhn argues that part of what happens in a scientific revolution is that the meanings of linguistic expressions change. So Kuhn would insist that the two incommensurable theories are not both expressed in English, rather, one is expressed in pre-revolutionary-English and the other in post-revolutionary-English and that these two languages are in fact not mutually translatable.

---

[357] Ibid., 184.

> Two men who perceive the same situation differently but nevertheless use the same vocabulary in its discussion must be using words differently. They speak, that is, from what I have called incommensurable viewpoints. How can they even hope to talk together much less to be persuasive.[358]

So the Kuhn-Feyerabend theory of incommensurability in science is not obviously inconsistent with Davidson's proposal. But what about Whorf's Linguistic Relativity Hypothesis? Davidson's proposal seems like an oversimplification of Whorf's view, for recall that Whorf held that both languages and people have conceptual schemes. According to Whorf, people inherit their schemes from the languages that they speak. Davidson, by the way, objects to this sort of proliferation of schemes.

> If conceptual schemes aren't associated with languages in this [Davidson's] way, the original problem is needlessly doubled, for then we would have to imagine the mind, with its ordinary categories, operating with a language with its organizing structure. Under these circumstances we would certainly want to ask who is to be master.[359]

Whorf would respond unhesitatingly: it is the conceptual scheme of language that is to be master over the schemes of individual minds. That is exactly what the Linguistic Relativity Hypothesis is all about. Consequently, Whorf would not agree with Davidson that the entire issue of incommensurability can be reduced to that of failure of translation. On the other hand, Whorf would agree that the schemes associated with languages are incommensurable if and only if the languages are not translatable. Indeed, failure of translation is precisely the phenomenon that Whorf repeatedly invokes to demonstrate that languages have incommensurable conceptual schemes associated with them.[360]

---

[358] T. S. Kuhn, The Structure of Scientific Revolutions, 2nd ed., (University of Chicago Press, 1970), 200.

[359] D. Davidson, op. cit., 184.

So Davidson is on the right track. If he can demonstrate that there is something wrong with the idea of translational failure, then there is something deeply wrong with the Linguistic Relativity Hypothesis, at least in its extreme forms. Davidson proceeds in two steps: (1) He argues that total failures of translation are impossible. (2) He argues that significant partial failures of translation are impossible.

Total Failure of Translation

All conceptual relativists distinguish between conceptual schemes and what the schemes are about, that is, the scheme-content distinction is essential to the relativistic doctrine. The basic idea is that there is an underlying, unchanging content that all schemes are directed at. This content is sometimes held to be "reality", sometimes "experience". The content is somehow ordered, organized or embellished by the conceptual scheme. Davidson insists that this scheme-content distinction needs further elaboration.

Davidson points out that authors like Whorf and Kuhn make frequent use of a number of metaphors that are intended to characterize the role of a conceptual scheme. Davidson quotes Whorf who says:

---

[360] See, for example, extensive discussion of a Nootka sentence in B. L. Whorf, Language, Thought and Reality, ed. by J. B. Carroll, (The M.I.T. Press, 1956), 242-43.

> ...language produces an organization of experience. We are inclined to think of language simply as a technique of expression, and not to realize that language first of all is a classification and arrangement of the stream of sensory experience which results in a certain world order...[361]

Davidson mentions other authors, Quine, for example, who speak of the conceptual organization of reality (rather than experience). What all these authors are getting at, Davidson says, is a supposed distinction between conceptual scheme and what it is about (reality or perhaps experience). According to proponents of the conceptual scheme idea, people confront reality (or experience) not directly, but through the mediation of their conceptual scheme. The scheme somehow organizes, structures, faces or fits reality (or experience). These metaphors can be understood in two ways according to Davidson.

(1) A scheme can be related to its content by organizing the content into objects and events, that is, by imposing an ontology onto the content. Since a scheme is best identified with a language (see above) we can see that this claim echos Quine's linguistic approach to ontological issues ('To be is to be the value of a variable'). In other words, we can make sense of the idea of a scheme organizing its content by attending to the referential structure of the language, in particular, its names, predicates, quantifiers and so on. Now, can different schemes organize reality (or experience) differently? The question becomes: can different languages have referential

---

[361] Ibid., 55.

structures that are somehow incommensurable. Davidson says the following:

> A language may contain simple predicates whose extensions are matched by no simple predicates, or even by any predicates at all, in some other language. What enables us to make this point in particular cases is an ontology common to the two languages, with concepts that individuate the same objects. We can be clear about breakdowns in translation when they are local enough, for a background of generally successful translation provides what is needed to make the failures intelligible. But we were after larger game: we wanted to make sense of there being a language we could not translate at all.[362]

The argument, then, is that the only way that a scheme (language) can organize its content is by imposing a referential structure upon it. But it is impossible to understand incommensurability in these terms, for even significant failures to translate the predicates of a language imply a common ontology that gives sense to the failure.

(2) The other way of understanding the metaphors offered by Whorf, Kuhn and others is to examine the way that whole sentences are related to content. That is, rather than attending to the referential apparatus of language, we attend only to whole sentences. But according to Davidson, sentences only relate to content in one simple way: either they are true or false. Claims that sentences somehow relate to reality in some other sense cannot be justified.

> The trouble is that the notion of fitting the totality of experience, like the notion of fitting the facts, or of being true to the facts, adds nothing intelligible to the simple concept of being true. To speak of sensory experience rather than the evidence, or just the facts, expresses a view about the source or

---

[362] D. Davidson, op. cit., 192.

> nature of evidence, but it does not add a new entity to the universe against which to test conceptual schemes. The totality of sensory evidence is what we want provided that it is all the evidence there is; and all the evidence there is just what it takes to make our sentences or theories true. Nothing, however, no thing, makes sentences and theories true; not experience, not surface irritations, not the world, can make a sentence true. That experience takes a certain course, that our skin is warmed or punctured, that the universe is finite, these facts, if we like to talk that way, make sentences and theories true. But this point is put better without mention of facts. The sentence 'My skin is warm' is true if and only if my skin is warm. Here there is no reference to a fact, a world, an experience, or a piece of evidence.[363]

So here is where we are in the argument: A conceptual scheme is supposed to be something that is somehow related to content. But Davidson has argued that when we try to make sense of this relation, all we end up with is the idea of the scheme being a set of sentences that are true, or at least largely true. Therefore, for conceptual relativity to be true, it must be the case that there can be two theories couched in different languages, both largely true, but not inter-translatable.

Now for the last step in Davidson's argument. He denies that two theories can both be largely true but not inter-translatable. He denies this because he believes our very understanding of the concept of truth is based on a prior understanding of the concept of translation. Therefore to claim that some foreign theory is largely true but not translatable is to say something incoherent.

Why does Davidson believe that the concept of truth is somehow linked to the concept of translation? Apparently because he endorses Tarski's method of defining truth for particular languages. Tarski's famous Convention T is a constraint on definitions of truth for particular languages. Convention T says that a truth definition of a language L must entail a sentence of the form:

---

[363] Ibid., 193-194.

s is true if and only if p

for every sentence of L, where 's' is replaced with a name of the L sentence, and 'p' is replaced by a translation of the L sentence into the metalanguage in which the truth definition is formulated.

So indeed, if you are a strict Tarskian you will be compelled by Davidson's argument, for you will not be able to make any sense of the notion of truth apart from the notion of understanding. What is amazing about Davidson's argument, however, is that in spite of what he says in "On the Very Idea of a Conceptual Scheme", he is not a strict Tarskian, and in fact in many of his other writings he explicitly rejects the idea that our grasp of truth depends on a prior understanding of translation or synonymy or any other related notion. Consider the following passage (part of which was already quoted in Chapter Two):

> Tarski's Convention T demands of a theory of truth that it put conditions on some predicate, say 'is true' such that all the sentence of a certain form are entailed by it. These are just those sentences with the familiar form '"Snow is white" is true if and only if snow is white'. For the formalized languages that Tarski talks about, T-sentences (as we may call these theorems) are known by their syntax, and this remains true even if the object language and metalanguage are different languages... But in radical interpretation a syntactical test of the truth of T-sentences would be worthless, since such a test would presuppose the

> understanding of the object language one hopes to gain. The reason is simple: the syntactical test is merely meant to formalize the relation of synonymy or translation, and this relation is taken as unproblematic in Tarski's work on truth. Our outlook inverts Tarski's: we want to achieve an understanding of meaning or translation by assuming a prior grasp of the concept of truth.[364]

If we agree with Davidson's 1973 "inversion" of Tarski, and I think we should for reasons given in Chapter Two, then his complex argument against the possibility of total failures in translation simply falls apart at the end.

But let us see if we can salvage Davidson's argument. Let us go back to the point where he concludes that the only sense that we can make of two conceptual schemes being incommensurable is that there are two theories (corresponding to the schemes), both largely true, but not inter-translatable. Davidson has failed to make the case that it is *conceptually* impossible that these theories could be true but not translatable, but perhaps as a matter of fact whenever two theories are largely true they will always be inter-translatable.

Although I suspect that this is the case, I have no idea how to develop an argument to prove the point. What I offer instead is an argument for a point that is somewhat weaker. My claim is that if there is a community of creatures that we recognize as having a conceptual scheme, then we will always be able to translate their scheme into ours. This is because (following Davidson) the only sense that we can make of the conceptual scheme idea is that there is a set of sentences that is largely true. Therefore to recognize a conceptual scheme is to recognize that there is a set of sentences that is largely true. But for us to recognize of a set of the creatures' sentences that they are largely true will require, in effect, that we formulate a set of T-sentences for the creatures' sentences, and this we will do in our language of course. But if we have done this then we have solved the translation problem, for the right hand side of the T-sentences will be a translation, in our language, of the creatures' sentences.

---

[364] D. Davidson, "Radical Interpretation", *Dialectica*, 1973, *27*, 313-328. Reprinted in D. Davidson, *Inquiries into Truth and Interpretation*, (Oxford University Press, 1984), 150.

This leaves open the possibility that there are creatures who in fact have conceptual schemes but which we do not recognize as having conceptual schemes. We have yet to establish an argument to rule out incommensurability in this case, although such an argument may be forthcoming. In any case, by building on Davidson's argument it has been shown that it is impossible that any tribe we recognize as having a conceptual scheme could have a scheme that is totally incommensurable with our own.

Partial Failure of Translation

Davidson also considers whether a somewhat watered-down version of conceptual relativity can be maintained. A lesser, although still remarkable, relativity would be evidenced if significant portions of a tribe's language could not be translated into our own.

Davidson argues that this is not possible either, because the principle of charity demands, as a constraint on radical interpretation, that we count the tribe members correct in most of their beliefs. Davidson's argument is focussed on showing that the principle of charity is, indeed, a necessary constraint on radical interpretation. However, in making his argument, Davidson focuses on his particular step-by-step approach to radical interpretation, in particular, the first step of identifying the sentences that Karl holds true. In what follows I will attempt to present the intent of Davidson's argument without relying on the specifics of Davidson's step-by-step method.

What the principle of charity rules out is the attribution of beliefs to a tribe that are massively different than our own. It does not rule out minor differences. Davidson's point is that we can only make sense of a difference of opinion against a background of shared opinion. The smaller the difference of opinion between two people, the more precisely it can be formulated. Wild differences of opinion quickly become obscure; if we try to attribute a difference of opinion that is too different from our own, we quickly are overcome with the feeling that we are not really sure what we are attributing. Consider the following examples:

(a) Karl believes that heavy objects fall faster than light objects.

(b) Karl believes that the sky is supported by pillars.

(c) Karl believes that books are sentient beings.

We have already ruled out the scenario that it is completely impossible to translate any part of Karl's conceptual scheme, so let us assume that we have already attributed to Karl a host of beliefs that are common to most people over eight years old in western societies. Karl believes that objects can hurt him if he walks into them; he believes that food is necessary for life; he believes that cars and airplanes are used for transportation; he believes that there are a great many people on earth; and so on. (This is the principle of charity in action.) Now, if we add to this stock of beliefs the additional belief (a), it will not cause any sort of crisis. Many people do, in fact, subscribe to this erroneous belief. As interpretors we have no doubts as to exactly what error Karl is subscribing to, for the erroneous belief, although perhaps inconsistent with some of Karl's other beliefs, is in some looser sense "compatible" with the standard beliefs of a normal twentieth century citizen of the Western world.

But consider attribution (b). To add some realism to the thought-experiment, assume Karl is your neighbor. You don't know him well, but have had numerous over-the-fence conversations with him and he seems like an ordinary sort of person. You know that Karl works as an accountant, that he and his family flew to Europe last year, and various other unremarkable facts about him and his life. Then someone tells you that Karl believes that the sky is supported by pillars. Your response (at least my response) will be, "What do you mean he believes that the sky is supported by pillars? The man is not an ancient Egyptian. How could he believe such a thing? What are you really trying to say about him?"

I don't deny that we could perhaps make sense of Karl having the belief that the sky is supported by pillars. But before we could accept such an attribution we would have to revisit a number of our previous belief attributions made in the name of the principle of charity. The principle of charity allows us to attribute to every western citizen over the age of eight years the belief that mankind has launched many missiles into space. This belief attribution is now problematic in Karl's case because of attribution (b). Does (b) mean that Karl believes that the sky is like the ceiling in a room, and if so, does Karl believe that the missiles sometimes smack into the ceiling, or that they haven't got that high yet, or that there really are no missiles (their existence has been fabricated by various government agencies)? There will be many questions of this type because (b) is not "compatible" with the general background beliefs that are warranted by the principle of charity.

Now consider (c). Once again this will be met with the response, "What are you talking about? How could he believe that?" And once again we can only make the attribution plausible by going back and revising Karl's general background beliefs. But now the revision will take on an eerie quality. If Karl believes that books are sentient beings, does he believe that rocks are sentient beings? What is the nature of this sentience; is it at all like our own? Do we have moral obligations to books (not moral obligations about books as important human artifacts, but moral obligations to books)? We may begin to doubt that we understand what Karl means by sentience at all, and that means that a component of the general shared beliefs previously attributed to Karl has just crumbled away. Obviously, the greater the number of attributions we make that are like (b) and (c), the more damage we do to the shared background beliefs. And if the shared background falls apart, we lose our grasp of attributions like (a) and eventually attributions like 'Karl believes that food is necessary for life' also fall apart. In other words, unless we assume that Karl and we share most beliefs, we cannot make any sense of our differences of opinion. If differences of opinion start to dominate over shared opinion, the whole process of interpretation - in which commonalities and disagreements are identified - starts to break down. If the breakdowns are severe enough then it simply becomes incoherent to try to make the attribution that is causing all the damage.

To complete the case against islands of incommensurability, consider the following attribution:

(d) Karl believes that p. (Where p is a proposition that cannot be expressed in our language.)

The claim that Karl has portions of his conceptual scheme that are incommensurable with ours is expressed in (d). However, (d) should be viewed as the extreme end of a continuum that starts with attributions like (a) through attributions like (c) and finally culminates in pseudo-attributions like (d). As the previous argument demonstrates, the further we get from the (a) end of the continuum, the more likely we are to destroy the common background of beliefs that give sense to the differences of opinion that we do have.

In claiming that (d) is at the end of the continuum I am refusing to accept that these supposedly incommensurable beliefs of Karl's can be isolated from Karl's other beliefs, the ones that we supposedly share with Karl. Karl's beliefs must have some sort of internal coherence (this, in fact, is part of what is demanded by the principle of rationality), but the "islands of incommensurability" thesis does not permit us to examine this coherence. However, by treating (d) as an extreme case of (c) we are, in effect, insisting on the right to try to grasp the content of Karl's obscure belief by examining its relation to Karl's other belief. If this examination results in a total breakdown of the interpretive exercise (which has been previously shown to be impossible) then so much the worse for the attribution of strange beliefs. The principle of charity is a necessary precondition for interpretation.

Conclusion

Donald Davidson has presented forceful arguments against the idea of total or partial incommensurability of conceptual schemes. I have suggested that there is an error in one of his arguments, but it can be rectified, at least in part. I have also modified another of his arguments so that it depends less on the specific step-by-step method of "holding-true". But Davidson's conclusions hold: extreme conceptual relativity is not possible.

As a sidelight, Davidson concludes that since there are no alternative conceptual schemes, there is no such thing as a conceptual scheme at all. Of course there are different languages, and each language has a relation to the world via truth-conditions, but we add nothing to this picture by saying that there is a conceptual scheme over and above the language. The "very idea of a conceptual scheme" is considered by Davidson to be the "third dogma of empiricism", to be rejected along with the analytic-synthetic distinction and epistemological reductionism, the two mistakes that Quine identified in his classic paper "Two Dogmas of Empiricism".

Davidson is happy to conclude that incommensurability is incoherent. However, there surely are differences between cultures, commensurable though they may be, and Davidson has provided no method of comparing and contrasting them (although, to be fair, this was not the problem that he set for himself). The next two sections examine some ways in which the language and thought of different groups can be compared and contrasted.

4.5 RELATIVITY OF REFERENCE

Bruce Aune has recently argued that Davidson has overstated the case against conceptual relativism.[365] Aune agrees with Davidson that the incommensurability thesis is not viable, and therefore extreme relativism (the idea that different people live in different "worlds" because of their conceptual schemes) makes no sense. However, Aune does think we can make sense of the notion of alternate conceptual schemes, even though these alternate schemes will be comparable from a more general perspective.

---

[365] B. Aune, "Conceptual Relativism", in J. E. Tomberlin, ed., Philosophical Perspectives 1: Metaphysics, (Ridgeview Publishing, 1987).

Davidson's argument against conceptual relativism was discussed in the last section, but can be summarized as follows.

1. Conceptual relativity is a rather vague thesis, but what it seems to come to, when clarified, is the claim that speakers share a conceptual scheme if they speak the same language or if they speak languages that are wholly translatable into each other. If they speak languages that are not translatable, or only partially translatable, then they have alternate conceptual schemes.

2. But total failure of translation is not possible. That is because each speaker's language is largely true (for being largely true is how we have to unpack the metaphors of a conceptual scheme "fitting" or "organizing" reality or experience). So conceptual relativity comes down to the idea of two languages being largely true but not translatable. And to hold this requires that we understand the concept of truth independently of the concept of translation. But reflection on Tarski's work shows that these concepts are not independent. Therefore the idea of total translational failure breaks down.

3. And partial failure of translation is not possible to any significant degree either. That is because in radical interpretation we have to solve for a vector of meaning and

belief, and the only way that we can understand the other as a "believer" is to understand him as being largely correct in his beliefs, and what this amounts to is viewing him as largely consistent with the beliefs that we hold true; therefore the principle of charity applies. This rules out the possibility of large areas of his beliefs (as conveyed by the meanings of his words) as being inscrutable to us; therefore if we can translate at all we must be able to translate (almost) everything (i.e., the principle of charity has to apply holistically).

Aune goes along with Davidson's first point for the sake of argument, but he does not find the second and third convincing. (I will not bother explaining in detail why Aune does not like Davidson's second and third points. Suffice to say Aune's criticism of the second point is obscure and the criticism of the third point is similar to the one I made against Davidson's linking of the concepts of truth and translation. However, I proposed an alternate argument for a conclusion that, although weaker than Davidson's, comes to much the same thing.) In any case, Aune thinks that given the Quinian view of translation that is largely accepted by philosophers, including Davidson, then Davidson does not need points two and three anyway. By "Quinian translation", Aune means that translation is viewed as the task of correlating the sentences of two languages, rather than correlation of the terms. On Quine's view, any speculation about the internal structure of sentences is intended only as a means to the primary goal, that of correlating sentences. Aune is correct in pointing out that this is a rather restrictive view of translation, since our ordinary concept of translation has it that the correlation of words is also a fundamental goal of translation. (How else do we explain the existence of inter-language dictionaries.) However, as Aune points out, Quinian translation is:

> .. a significantly weak relation - so weak that a theoretically acceptable translation manual will no doubt exist for any two languages or dialects, at least if they are comparably rich.[366]

So if Quine's view on translation is right, and if point 1 above is right, then it follows that alternate conceptual schemes are impossible.

Aune on Translation

But Aune does not accept Quine's weak conception of translation. He thinks that a much stronger conception of translation can be derived from Goodman's 1949 essay "On Likeness of Meaning".

> Goodman's criterion [for measuring likeness of meaning between referential terms] is based on the distinction between the primary and secondary extension of a term of a predicate. The denotation of a term is its primary extension; its secondary extension is the denotation of compounds in which it occurs. Goodman introduced this distinction to point out a clear extensional difference between terms that have, like 'centaur' and 'unicorn', the same extension but differ in meaning.[367]

Using this criterion of synonymy it is possible to distinguish two types of translation.

---

[366] Ibid., 270.

[367] Ibid., 275.

1. "Strong Translation" - which is what we get when we apply Goodman's criterion on a cross-language basis. But note that if the languages are not historically related then it may not be possible to pull this off simply because no pairs of referential terms agree on secondary or even primary extension.

2. "Weak Translation" or "Crude paraphrase" - which is what we get when we set up a Quinian translation manual between two languages that are not translatable in the strong sense characterized above.

Aune says that we can use this distinction to sympathetically interpret writers like Whorf. Whorf can be interpreted as claiming that Nootka and English are not strongly translatable. In making his case Whorf gives a number of crude-paraphrases. But he does not thereby contradict himself.

Contrast this with Davidson's interpretation of Whorf:

> Whorf, wanting to demonstrate that Hopi incorporates a metaphysics so alien to ours that Hopi and English cannot, as he puts it, be calibrated, uses English to convey the contents of sample Hopi sentences.... The dominant metaphor of conceptual relativism, that of different points of view, seems to betray an underlying paradox. Different points of view make sense, but only if there is a common coordinate system on which to plot them; yet the existence of a common coordinate system seems to belie the claim of dramatic incomparability.[368]

Aune feels that Davidson sees Whorf as paradoxical only because Davidson does not recognize that there are two standards for translation. Aune has no problems agreeing with the Whorfian possibility that two languages might not be strongly translatable, and consequently "those languages cannot involve anything that is reasonably called a common conceptual scheme",[369] even though we might try to convey some of the flavor of the incommensurability by means of a weak translation.

---

[368] D. Davidson, "On the Very Idea of a Conceptual Scheme", Proceedings and Addresses of the American Philosophical Association, 1974, 47. Reprinted in D. Davidson, Inquiries into Truth and Interpretation, (Oxford University Press, 1984), 184.

Aune on Conceptual Schemes

Unlike Davidson, Aune believes in conceptual schemes, but he does not associate them one for one with languages the way Davidson does in point 1 above. (Davidson, of course, only establishes this correlation to knock it down, since he doesn't believe in conceptual schemes at all.) Rather, Aune thinks that several conceptual schemes can be associated with any language.

How do schemes get associated with language? Well the 17th and 18th century rationalists and empiricists argued, in essence, that schemes are the outcome of our innate mental contents and/or capacities. Strawson, more recently, argues that conceptual schemes are based on analytic relations between primitive and derived terms.

Aune accepts none of these arguments. He argues, in a somewhat pragmatic vein, that language is just a system of verbal behavior with no inherent conceptual system. However, throughout the intellectual history of a speech-community, different thinkers can impose their conceptual schemes on a language. Aune discusses the metaphysics of Leibniz and Russell as two competing schemes that have been "imposed upon" Indo-European.

---

[369] B. Aune, op. cit., 279.

(The idea of multiple conceptual schemes within a single language is, in fact, consistent with Whorf, who talks about Newtonian and relativistic physics as two different conceptual schemes. But Whorf differs from Aune in stressing the unconscious nature of conceptual schemes, whereas Aune seems to suggest that they get introduced very consciously, and they can be historically identified as occurring at a certain time as the product of a particular thinker. In one of his papers, Whorf hypothesized that Newton got his idea of space-time from the conceptual scheme of Indo-European. But note that if Whorf wants to claim that Einstein also derived his alternate conceptual scheme from Indo-European, then Whorf will have to countenance the idea that there are multiple unconscious conceptual schemes associated with Indo-European. I somehow doubt that he would want to say this. The idea of multiple conceptual schemes fits much more happily with Aune's idea that they are consciously introduced, than with Whorf's idea that they are generated unconsciously.)

Aune is not at all clear about how an individual might go about imposing his scheme on a language. Toward the end of the essay he says this:

> The only way a mind ... can organize a world is by representing it in an organized way. What we construct is a world picture - or better, a world story, for the picture consists of judgments, and these, when asserted, yield stories rather than pictures or objects in a world.[370]

---

[370] Ibid., 285.

Not only is this lacking in detail, but it seems to contradict Aune's focus on the referential structure of a language. For if a thinker imposes a scheme by telling a story (i.e., asserting a bunch of sentences) then it seems that we are being asked to interpret that speaker on a sentence-by-sentence basis, rather than a term-by-term basis. And this leads to all the Quine-Davidson problems with inscrutability of reference, with the inevitable conclusion that a referential structure is merely an under-determined instrumental device of the interpreter. In fact, Aune's last sentence in the quotation above seems to suggest a "posit" approach to objects. But what this does, it seems to me, is to totally undermine Aune's reliance, in his characterization of conceptual schemes, on Goodmanish considerations about referential structures. As further evidence of this contradiction compare the passage quoted above with this next one, and notice that this next one says that it is referential structure, not sentential "stories", that organize the world:

> A conceptual scheme can organize a world by providing a system of predicates that apply to and thereby classify (or systematically interrelate) objects discernible in it; it may fit a world in the sense that the relevant predicates are satisfied (perhaps to an adequate degree) by the objects thus discernible. In taking objects as the correlates of fitting, I betray my conviction that facts are really fictions (not real things.) And in contrast to Davidson, I believe that real things [i.e. objects, D.M.] do make statements true.[371]

So what Aune needs, if he is to make his case for alternate conceptual schemes in a language, is a better theory of what it is for a thinker to impose a new scheme on a language, one that is consistent with his view of conceptual schemes as object classification systems.

---

[371] Ibid., 284.

Summary of Aune's Argument

1. Conceptual schemes are referential structures that can be used to "distinguish different objects in the world and classify what they distinguish in different ways." (p. 285)

2. Conceptual schemes are not, as Whorf claims and Davidson assumes (in setting up his straw man argument), inherent in a language. Rather, they have to be imposed on a language by particular thinkers. (Aune is not too clear on the details of this.) Furthermore, conceptual schemes are not subliminal in the sense suggested by Whorf. Conceptual schemes have to be introduced into a culture by individuals who are consciously intending to introduce them. Thus all conceptual changes will have a definite history (even though it may be quite difficult to reconstruct this history in some cases).

3. It is possible, therefore, that several conceptual schemes can be expressed within a single language.

4. Conceptual schemes can be compared by using Goodman's method for determining likeness of meaning. Although Goodman was interested in working within a single language, his method can be extended across languages, and this yields a notion of "strong" term-by-term translation that can be contrasted with Quine's "weak" sentence-by-sentence notion of translation.

5. Conceptual relativism can be characterized as the claim that there are schemes (referential structures) that are not isomorphic in Goodman's sense. Conceptual relativism is therefore a viable empirical hypothesis, and may well be true for the language pairs discussed by Whorf. However, this new version of the Linguistic Relativity Hypothesis based on referential structures is not compatible with the incommensurability thesis.

6. Davidson's refutation of the incommensurability thesis is not objectionable to Aune, but Davidson's rejection of "the very idea of a conceptual scheme" is not satisfactory because it recognizes only the weak notion of translation. Furthermore, Davidson is unfair to Whorf. Davidson says that Whorf contradicts himself by attempting to provide translations between languages claimed to be incommensurable. Aune points out that given the distinction between strong and weak translation, it is possible to give Whorf a much more sympathetic reading.

Comments on Aune's Argument

As I see it, there are two problems with Aune's argument. First, he has failed to provide an adequate account of how an individual member of a speech community goes about "imposing" his conceptual scheme on a language. This process is critical to Aune's theory of conceptual schemes, but the few suggestions that he provides seem more consistent with a Davidsonian sentence-level view of our intellectual architecture, rather than a term-level architecture that Aune requires.

Secondly, Aune assumes that the notion of a term-level intellectual architecture makes sense. Now Aune is not suggesting that return to the "concept" concept as our basic element of intellectual attribution. He would presumably agree with the arguments of Chapter Three that sub-propositional elements make sense only in the context of propositions. However, some authors, notably Davidson, have recently argued that sub-propositional elements have are merely theoretical constructs in our semantic theories, and that they have no reality, psychological or otherwise, at all. Furthermore, there is a profound indeterminacy associated with the referential component of a semantic theory. The referential component can be constructed many different ways, even though all the ways are equally compatible with all the empirical evidence we might have about the linguistic behavior of a speech-community. For Davidson, only the truth-condition, as realized in a T-sentence, has any sort of real empirical content.

I am going to ignore the first problem with Aune's paper. Let us assume for now that it is possible to fabricate some sort of theory about how referential structures are introduced; perhaps the much-discussed "causal theory of reference" will provide the foundation of such a theory. However, the second problem cannot be avoided. If, as Davidson holds, referential structures are merely instrumental, then Aune's theory of conceptual schemes will fall apart. Consequently, the discussion will now turn to Davidson's thesis.

Davidson on Reference

In his paper, "Reality Without Reference", Davidson points out that the concept of reference seems to pose a dilemma for the theorist of language who shares the view that a theory of linguistic action must incorporate something like a Tarskian truth theory.[372] By "reference" Davidson means a relation between names and what they name, complex singular terms and what they denote, one-place predicates and what they are true of, two-place predicates and the pairs they are true of, and so on. We may say that a "referential structure" of a language is given by those clauses of a Tarskian truth theory that give the denotations of names and complex singular terms, and the satisfaction conditions of predicates.

Davidson points out that on the one hand, it seems that reference is an essential concept, since a referential structure is an essential component of every Tarskian truth theory. On the other hand, a Tarskian theory, in itself, does not give an explication of the concept of reference, rather, the concept of reference is simply assumed in the theory, and is employed by the theory in that the theory will include a list of clauses that fix the reference of names, singular terms and predicates.[373]

---

[372] D. Davidson, "Reality Without Reference", Dialectica, 1977, 31, 247-253. Reprinted in D. Davidson, Inquiries into Truth and Interpretation, (Oxford University Press, 1984).

[373] In a classic article, Hartry Field clearly points out the inadequacy of Tarski's approach as an explication of the concept of reference.Hartry Field's classic article. H. Field, "Tarski's Theory of Truth", Journal of Philosophy, 1972, 69, 347-375.

The fact that a Tarskian truth theory uses the concept of reference, but does not explicate it, has led many writers to claim that Tarskian truth theories must be supplemented with a theory of reference in order to yield an adequate account of natural language. But Davidson says that this will lead us down the wrong road: to a "building-block" theory of language, in which the strategy is to empirically identify the referential roles of words independently of their roles in sentences, and then later to some explicate how words somehow combine into sentences. Davidson's position is that this is a hopeless strategy, since words have no roles, referential or otherwise, independently of their roles in sentences. (The reasoning behind this claim is pretty much the same as that which I used in Chapter Three to point out that the concept of a concept cannot be explicated independently of the role that concepts in propositional attitudes.)

This leads Davidson to point out that reference leads to the following paradox, for which he proposes a resolution.

> Here then, in brief, is the paradox of reference: There are two approaches to the theory of meaning, the building-block method, which starts with the simple and builds up, and the holistic method, which starts with the complex (sentences, at any rate) and abstracts out the parts. The first method would be fine if we could give a non-linguistic characterization of reference, but of this there seems no chance. The second begins at the point (sentences) where we can hope to connect language with behavior described in non-linguistic terms. But it seems incapable of giving a complete account of the semantic features of the parts of sentences, and without such an account we are apparently unable to explain truth.
>
> To return to the central dilemma: here is how I think it can be resolved. I propose to defend a version of the holistic approach, and urge that we must

> give up the concept of reference as basic to an empirical theory of language.[374]

Davidson's proposal is to relegate the concept of reference to a merely instrumental role. A theory of language gets it empirical content only at the sentential level, according to Davidson. As we saw in section 4.3, Davidson's approach to radical interpretation is to fix the truth-conditions of sentences on the basis of observing what utterances are held true by particular individuals under particular circumstances. This is all the evidence there is in constructing a theory of linguistic behavior. However, in constructing a theory of language we have to go beyond the finite evidence available to us at any given time, and construct a theory that allows us to interpret an unlimited number of potential utterances. (This is Lewis' principle of generativity in action.) To do this we construct a Tarskian theory, which in turn requires the postulation of a referential structure. Davidson claims that given a body of evidence regarding utterances held true in a speech community, there will be an unlimited number of alternative referential structures that will generate any particular infinite set of T-sentences. Consequently, to adopt one of these referential structures over another is merely a matter of theoretical convenience: there is nothing empirical at stake. For Davidson speech acts are real, and therefore so are sentences; words, on the other hand, are a fiction of the theoretician.

---

[374] D. Davidson, _op. cit._, 221.

If Davidson is correct in his "instrumental" view of referential structures, then Aune's theory of conceptual relativity will fall apart, for it depends on the notion of reference as a real feature of language. Aune's view is that conceptual schemes are introduced to speech communities by individuals who propose new referential structures for words. Davidson denies that referential structures have any extra-theoretical reality, therefore they cannot be "introduced" into a speech community, nor can they be invoked to support a theory of conceptual relativity.

The Causal Theory of Reference

In the last two decades a number of philosophers of language have developed a new theory of reference, commonly known as the causal theory of reference.[375] The theory is "new" in that it contrasts with what has come to be known as the "traditional" theory of reference; actually a loose collection of doctrines that is common to a wide range of theoreticians, including the British empiricists (Locke, Berkeley, and Hume), the logical positivists of the twentieth century, the later Wittgenstein, and contemporary philosophers of language such as John Searle.

Both the traditional theory and the new theory hold that reference is real. However, the referential realism of the traditional theory quickly succumbs to Davidson's concerns about how the impossibility of identifying the semantic roles of words independently of their contribution to the semantic role of sentences. On the other hand the causal theory, as we shall see, does suggest that referential structures have a role that is not exhausted by their contribution to the generation of T-sentences within a Tarskian truth theory.

---

[375] For an introductory account, see S. Schwartz, "Introduction", in S. Schwartz, ed., Naming, Necessity and Natural Kinds, (Cornell University Press, 1977).

Let us examine this point more carefully. For Davidson there is only one constraint on how the theorist should construct referential structures. That constraint is that the referential structure must generate T-sentences such that when those T-sentences are worked into a theory of speech acts, the overall linguistic theory is consistent with the evidence regarding the utterances that are held true in a speech community by particular speakers under particular circumstances. Because this single constraint allows a great deal of indeterminacy in the selection of referential structures, Davidson concludes that "we must give up the concept of reference as basic to an empirical theory of language"; that is, reference should be viewed instrumentally rather than realistically. If Davidson is correct, then Aune's attempt to rebuild the notions of conceptual schemes and conceptual relativity on the basis of referential structures is doomed to failure.

However, the advocates of the causal theory of reference hold that there is a second constraint on the theorist's development of referential structures. The causal theory of reference allows us to establish connections between particular words and objects (in a manner to be explained below). Thus any referential structure that the theorist proposes must also be consistent with the causal theory of reference. An advocate of the causal theory need not disagree with Davidson's point that in gathering evidence we must start with speech acts, which are always the utterances of whole sentences. However, the causal theorist would argue this does not mean that the truth-conditions attributed to those sentences are somehow real while the referential structures are somehow not. For example, the causal theorist Michael Devitt writes:

> I have agreed that the evidence for our explanation is at the level of sentences because the sounds that prompt our investigations are sentences. And truth is a property of sentences and reference a property of words. However, I do not understand the link between this and the mysterious claim that truth is, but reference is not, a place of 'direct contact' with the non-linguistic world. The whole semantic theory including talk of truth and reference is tested at once by the evidence, and rather indirectly tested at that, for there are many other theories involved as well.[376]

Davidson holds that one garners empirical support for a semantic theory of language by studying speech acts and reading off the truth-conditions of sentences fairly directly from those speech-acts. The causal theorist, such as Devitt, holds that one garners support for a semantic theory by studying speech acts and then rather indirectly, one simultaneously reads off both the truth-conditions of sentences and the referential relations between words and parts of the world. One is guided in the second task by the causal theory of reference.

Because the causal theorist will not be faced with the degree of indeterminacy regarding the empirical identification of referential structures that Davidson faces, and because the causal theorist has a theory about reference that elevates it from merely instrumental status to a living, breathing part of linguistic life, the causal theorist has a thoroughly realistic conception of reference. It would seem, then, that if the causal theory is correct then referential structures are real, and can therefore be used to support Aune's theory of conceptual schemes. The questions we must know ask are: what is the causal theory of reference, and is it true?

---

[376] M. Devitt, Designation, (Columbia University Press, 1981), 122.

The "traditional" theory of reference holds that every referring term (name, complex singular term, or n-place predicate) has an intension and an extension. An extension is the object referred to by a name or singular term, or a set of objects (or n-tuples of objects) that satisfy a predicate. Extensions are determined by intensions. In early versions of the traditional theory, intensions were thought of as attributes. Thus the term 'chordate' has as its intension the attribute of having a heart. This intension determines the extension which is a set containing, among other things, you and me. More recent versions of the traditional theory hold that intensions (at least of singular referring terms) are determined by definitions, thus the intension of 'Aristotle' is given by a conjunction of defining descriptions such as 'the teacher of Alexander', and so on. A variant on this (advocated by Wittgenstein and Searle) is to say that the intension is given by a cluster of defining descriptions, some of which are perhaps false.

The causal theory denies that the extensions of terms are determined by their intensions. Part of the causal theorist's argument is to show that there are grave problems with the traditional theory. For example, although the cluster version of the traditional theory allows that some of our defining descriptions of 'Aristotle' may be false, the cluster theory leads to unacceptable results if it turns out that all our descriptions are false. Perhaps there was a man named Aristotle who was neither a teacher of Alexander nor a student of Plato nor a philosopher, but that these things began to be believed of Aristotle, and that these beliefs have been perpetuated to the present day. Now, if these mistaken beliefs were all to come to light, then the traditional theorist would have to conclude that 'Aristotle' does not refer, and a sentence such as 'Aristotle once visited Sicily' would have no truth value. However, given supposition that there was a real man named Aristotle about whom a number of false beliefs came to be perpetuated, the traditionalist theory rings false. The sentence 'Aristotle' does refer and the sentence 'Aristotle once visited Sicily' does have a truth-value, and therefore the traditional theory must be false. (The causal theorists have developed a great number of such counter-examples to the traditional theory of reference.)

If intensions do not determine the extensions of terms, what does? Davidson would say that nothing does other than the arbitrary decisions of the theorist of language; terms don't "really" have extensions. The causal theorists, on the other hand, do believe that terms really have extensions, and that these are determined by initial acts of "baptism", which are a type of speech act in which names are given to things; in other words, ostensive definitions. A speaker using the name, perhaps hundreds of years later, succeeds in referring if there is a causal chain that somehow connects his usage of the name to the original baptism. (The exact nature of this causal relation has been a subject of much discussion in the causal theory of reference literature.) Furthermore, the speaker will succeed in referring if most, or even all, of his beliefs about the object referred to are false.

The causal theorists eliminate the need for extension-determining intensions. Names (and singular terms and predicates) can stand in referential relations to things (or sets of things) provided that there are appropriate causal chains between utterances of the names and initial "baptisms" in which the name was ostensively defined (or, in the case of predicates, a "prototype" of the predicate was ostensively defined).

It is my view that some version of the causal theory is probably true, however, to try to make this case would take us too far afield. Assuming that the theory is true, then, what is its significance to the Linguistic Relativity Hypothesis?

The significance is that the causal theory points to a body of historical evidence regarding our speech community; evidence about baptisms and causal chains that lead from baptisms to speech acts in which the referring terms are used. If we are in the business of developing a Tarskian style theory of language for the speech community, this evidence will act as a constraint on the types of referential structures that we postulate within our Tarskian theory. This is a constraint that Davidson has failed to recognize, and therefore by adopting the causal theory and its constraints we avoid Davidson's instrumental attitude towards referential structures. With a realistic attitude toward referential structures we are able to adopt Aune's approach to conceptual schemes, viz., that schemes are associated with referential structures. This allows for the possibility of the type of conceptual relativity that Aune describes in his article (which, as we have noted, varies significantly from Whorf's version of conceptual relativity).

The Inscrutability of Reference

The idea of using the causal theory of reference as a constraint on the selection of referential structures within our overall Tarskian theory of a language has been explicitly endorsed by causal theory advocates such as Michael Devitt. Their line of reasoning was sketched in the previous sub-section. However this line of reasoning is not without its problems. Both friends of reference[377] and foes of reference[378] have argued that the causal theory of reference does not help in reducing the level of indeterminacy involved in selecting a referential structure within a Tarskian theory.

---

[377] H. Field, "Conventionalism and Instrumentalism in Semantics", Noûs, 1975, 9, 375-405.

[378] D. Davidson, "The Inscrutability of Reference", The Southwestern Journal of Philosophy, 1979, 10, 7-19. Reprinted in D. Davidson, Inquiries into Truth and Interpretation, (Oxford University Press, 1984).

The causal theory does not help because of the inscrutability of reference, an issue that was first raised by Quine.[379] Recently, it has become common to explain the thesis of the inscrutability of reference using the idea of a permutation of the universe. I will do the same in what follows.

Suppose that for some language L there is a reference scheme R such that for every name of L, R assigns the object to which the name refers, and for every predicate of L, R assigns the set of objects that satisfy the predicate. (Assume that there are no complex singular terms in L, and that all predicates are one-place.) Suppose further that P is a mapping of every space-time point in the universe onto another point one kilometer away in some fixed direction. If we treat physical objects as sets of space-time points, and if O is an object characterized by the set

{x ¦ x is a space-time point and a constituent of O}

then P will define a new object Op characterized by the set

{P(x) ¦ x is a space-time point and a constituent of O}.

Clearly, the permutation P defines a total one to one function from the normal objects of the universe to a set of permuted objects. (Obviously, these new "objects" will not have the crisp natural boundaries that we usually associate with physical objects.)

This permutation of objects can now be used to define an alternate reference scheme for L. The new reference scheme Rp assigns to each name the object P(y), where y is the object assigned to that name by R. The new reference scheme Rp assigns to each predicate the set

---

[379] W. V. O. Quine, Word and Object, (The M.I.T. Press, 1960), chapter 2; and more fully in the title essay in Ontological Relativity and Other Essays, (Columbia Univeristy Press, 1969).

{P(y) ¦ y satisfies the predicate under R}.

Now, note that if L is reinterpreted according to the new reference scheme, the truth-conditions of all the sentences will be the same. Under the normal reference scheme R, we would have the following:

(1) 'Fido is asleep' iff Fido is a member of {x ¦ x is asleep},

whereas under the reference scheme Rp we would have:

(2) 'Fido is asleep' iff P(Fido) is a member of {P(x) ¦ x is asleep}.

Now, note that because of the way that P has been defined, it will be the case that:

(3) Fido is a member of {x ¦ x is asleep} iff P(Fido) is a member of {P(x) ¦ x is asleep}.

From (3) we can see that the right hand side of (2) is logically equivalent to the right hand side of (1). In other words, 'Fido is asleep' will have the same truth conditions no matter if we use the original reference scheme R or the permuted reference scheme Rp. This will be true of every sentence of L. Moreover, there are endless numbers of permutations of the universe that will allow us to generate endless numbers of alternate reference schemes, all of which will have the same truth conditions for the sentences of L. The thesis of "the inscrutability of reference" is that we have no good reasons for selecting R over Rp or any of the other alternate reference schemes. More precisely, the physical evidence regarding the behavior of the members of the speech community is equally compatible with all these different reference schemes.

The hope of the causal theorist is that the causal theory will allow us to select R as the "correct" reference scheme over all the competitors, thus providing a way to defeat the inscrutability of reference. The causal theory allows us to establish a direct relationship between 'Fido' and Fido, one based on a baptism and a causal history involving the use of 'Fido' within the speech community. This relationship will not hold between 'Fido' and P(Fido).

However, these hopes are dashed when we consider that if there is a causal theory C that establishes a connection between 'Fido' and Fido, then there is another theory, Cp, that establishes a connection between 'Fido' and P(Fido). Cp is an alternate causal theory that says 'Fido' denotes x if and only if someone once baptized $P^{-1}(x)$, and that there is a causal history of the use of 'Fido' extending back to that baptism. Therefore, on the Cp theory, 'Fido' will denote P(Fido). And there is nothing in the physical evidence of the speech-community's behavior that will justify a preference of Cp over C.

It might be countered that the Cp theory is patently absurd. Why would the members of a speech-community engage in a life-long intercourse (verbal and otherwise) with a set of objects, while really referring to a permuted set of "objects" one mile away? We seemed to be faced with a situation where the physical evidence supports causal theories C and Cp equally well. However, the sheer absurdity of Cp should convince us that we simply do not properly understand the way in which physical evidence can support a referential scheme, for relation between evidence and referential schemes has to be such as to rule out theories like Cp.[380]

So perhaps our attitude should be that the reference scheme R that is supported by the causal theory C is to be "obviously" preferred over the reference scheme Rp that is supported by the causal theory Cp. If the physical evidence seems to support both theories equally well then we should sadly conclude that we obviously don't understand the evidential relation adequately at this time.

The problem with this approach is that there are permutations of the universe that produce alternate reference schemes supported by alternate causal theories, but these alternate schemes and theories are not patently absurd. Quine has provided a wealth of such examples. Consider a permutation of the universe that, in effect, maps each object onto the set of undetached parts of that object. Or consider the permutation that maps each object onto the set of stages of that object. Each of these permutations will support an alternate reference scheme and an alternate causal theory. However, these alternate reference schemes are not "obviously absurd". Therefore we cannot appeal to our intuition to rule out the doctrine of the inscrutability of reference.

---

[380] Any argument to a strange conclusion can be turned into a reductio ad absurdum against one or more of the premises. The approach taken in this paragraph is similar to Nelson Goodman's in his classic Fact, Fiction and Forecast, 4th ed., (Harvard University Press, 1983). Goodman argued that our inductive logic, as currently understood, supports generalizations using the "odd" predicates 'grue' and 'bleen' just as well as it supports generalizations using the "normal" predicates 'blue' and 'green'. Goodman's real point is not that the alternate predicates are just as good, but rather, that there is something defective about our understanding of induction that allows such a conclusion.

Here is another way of looking at the problem. Crudely stated, the causal theory of reference says that for a name to refer to an object, someone once must have ostensively defined that object by, say, pointing to it and uttering 'Fido'. But what object was he pointing to: Fido, the set of undetached parts of Fido, the temporal stages of Fido, etc.? All these options, and more, are compatible with the evidence, and all produce different schemes of reference.

The Reality of Reference - A Defence

Here is a summary of where we are:

1. Bruce Aune has argued that the idea of conceptual schemes, and a limited degree of conceptual relativity, is valid. Aune wants to base his idea of conceptual relativity on the reference schemes of a language.

2. However, Donald Davidson has argued that reference schemes are not "real". That is, the only evidence that we can gather about linguistic behavior in a speech-community is at the level of sentences. True, as theorists of language we have to include a reference scheme in our Tarskian theory of the language, but the reference scheme should simply be viewed as an arbitrary and instrumental aspect of the theory. Reference has no independent reality. Clearly, if Davidson's ideas on

reference are correct then Aune is on the wrong track; for how can we base a theory of conceptual schemes (presumably intended to be about psychological reality) on the concept of reference, if reference is merely a theoretician's fiction?

3. Over the past two decades a new theory of reference has emerged. The so-called "causal theory of reference" rejects the traditional idea of "intensions" determining "extensions", and replaces it with a socio-historical account of why terms have the extensions they do. A key part of this theory is that terms get their extensions by human activities of "baptizing". So presumably, by studying the acts of "baptism", or ostensive definition, we can gather direct evidence about relation between names and predicates and their extensions. Therefore it appears that Davidson's instrumentalist view of reference is wrong, and the way is clear for Aune to develop his theory of conceptual schemes.

4. But not yet. The thesis of the inscrutability of reference states that the physical evidence will be consistent with infinitely many alternate reference schemes, even if we take causal considerations into account. Perhaps some of the more outlandish alternative reference schemes can be ruled out simply on the grounds of sheer implausibility (the idea being that there must be some inductive principle that we don't yet know, but which will rule them out), but there are still many significantly different alternative reference schemes that are more or less equally plausible. So once again, the situation looks bad for Aune's theory.

My view is that it is really not so bad. I agree that Quine is correct in saying that 'rabbit' might just as well refer to undetached rabbit parts as to rabbits. So there is a kind of inscrutability when we try to read off linguistic facts from the physical facts. However, this inscrutability is better understood as "inscrutability of the language used" rather than "inscrutability of reference". To appreciate my point, consider a language from a purely formal point of view. Let us take a formal language to be an ordered pair, the first element being an infinite set of T-sentences, one for each sentence in the language, and the second element being a Tarskian truth definition that generates all the T-sentences. Let X be the set of all formal languages. Now, any natural language, say English, can be construed as a subset of X, where each member of the subset is consistent with the physical evidence regarding the English speech-community. So let us say that English can be construed as the set E, which is a subset of X such that each member of E is compatible with the physical evidence. Now, some of the membership of E will be explainable through the doctrine of inscrutability of reference. But not all will be. There can also be a doctrine of the "inscrutability of grammar" where, by 'grammar', I mean a Tarskian truth definition minus the reference scheme. Clearly, E will contain many alternate grammars, all of which are compatible with the evidence.

In fairness then, we should speak of the inscrutability of language in total, rather than narrowly focusing on the inscrutability of reference. The inscrutability of language means simply that the physical evidence will never allow us to select one formal language as the language used by a speech-community, but rather, there will always be a set of compatible formal languages.

Reference is no less subject to this inscrutability than other aspects of linguistic description, but neither is it any more subject to it. Consequently, reference is just as "real" as any other aspect of linguistic description, including the T-sentences themselves. (Any finite body of physical evidence will be compatible with alternate sets of T-sentences.) The theorist's discussions of reference should not be put into a special epistemological category. When trying to determine what members of X should be included in E, on the basis of the physically described behavior of the English speech-community, we are involved in a holistic enterprise. Alternate reference schemes and grammars will be played off against one another, and the entire linguistic description will be played off against intentional attributions, i.e., the beliefs and desires that are ascribed to the members of the speech community. No part of this overall interpretation has any more or any less "reality" than any other part.

Davidson usually promotes a holistic approach to radical interpretation, however, as we have seen, he seems to want to put reference into a special epistemological category. However, he does not seem to be able to consistently hold that position, for in the following passage he reverts back to a thorough holism:

> There is, then, a reasonable way to relativize truth and reference [my emphasis]: sentences are true,

> and words refer, relative to a language.... [I]t is not entirely an empirical question what [formal] language a person speaks; the evidence allows us some choice in languages, even to the point of allowing us to assign conflicting truth conditions to the same sentence. But even if we consider truth invariant, we can suit the evidence by various ways of matching words and objects. The best way of announcing the way we have chosen is by naming the language; but then we must characterize the language as one for which reference, satisfaction, and truth have been assigned specific roles. An empirical question remains, to be sure: is this language one that the evidence allows us to attribute to this speaker?
>
> What permits us to choose among various languages for a speaker is the fact that the evidence - attitudes or actions directed to sentences or utterances - bears not only on the interpretation of speech but also on the attribution of belief, wants and intentions (and no doubt other attitudes). The evidence allows us a choice among languages because we can balance any given choice by an appropriate choice of beliefs and other attitudes. This suggest one more way we could relativize a theory of truth or reference: given certain assumptions about the nature of beliefs and other attitudes, we could show that, once we have decided what a person's attitudes are, the choice of a language is no longer up for grabs. Given a comprehensive account of belief, desire, intention and the like, it is an empirical question what language a person speaks. And so we have, at last, a rather surprising way of making significant sense of the question what a word refers to.[381]

Let us distinguish two types of instrumentalism. "Local instrumentalism" holds that some local component of a theory is a fiction of the theoretician. "Holistic instrumentalism" holds that there are an number of alternate theories that suit the evidence, although each component of the alternate theories has a "realistic" interpretation once the decision has been made to go with that theoretical alternative. Davidson seems to have presented two types of instrumentalism regarding reference. First he argued that reference is an example of "local instrumentalism" within our overall linguistic theory. In the passage just quoted he seems to be arguing for a "holistic instrumentalism" as regards our overall theory of interpretation, with no special consideration being given to reference. I believe that Davidson's local instrumentalism is on the wrong track. What seems much more plausible is the idea of holistic instrumentalism, or to say the same thing in other words: the total interpretation of a speech-community (including the ascription of a language and propositional attitudes) is underdetermined by the evidence.

---

[381] D. Davidson, op. cit., 240.

The issue being dealt with is: are reference schemes real? The answer is that they are real, relative to a formal language, and a natural language is best construed as the set of formal languages that is consistent with the physical evidence. Assume that it is possible to describe this set abstractly, that is, rather than enumerating all the formal languages that are compatible with the evidence, assume that we are capable of finitely stating some properties (in linguistic terms) that characterize exactly those formal languages in set E. Presumably, then, there will also be a way of abstractly characterizing the reference schemes that are compatible with the evidence. (The abstract characterization will have to be wide enough to encompass the following sorts of alternatives: 'rabbit' refers to rabbits, to undetached-rabbit-parts, to rabbit-stages, and so on.) This abstract characterization of the set of reference schemes consistent with the evidence can be called the 'referential structure' of the natural language.

Now, that we are armed with the notion of a 'referential structure' for a natural language, let us return to Aune's theory of conceptual schemes. If we simply assume that Aune's theory of conceptual schemes is based on referential structures, we are exempt from any concerns about the inscrutability of reference. It seems perfectly clear that a Newton or an Einstein might introduce a new referential structure (isn't that what they did?) and in Aune's sense, thereby introduce a new conceptual scheme.

As noted above, the details of Aune's theory are far from worked out. He has not stated the details of how these new conceptual schemes (referential structures) get introduced. However, assuming that such details can be worked out, we can see that a version of conceptual relativity is possible. But note that it is a theory of conceptual relativity that implies neither the incommensurability thesis nor Whorf's thesis. For on Aune's view it is people that cause conceptual variation, not language. Language is merely the tool that conceptual innovators use to express their new ways of looking at things.

4.6 RELATIVITY OF REASONING

It would be possible for me to conclude this work at this point. My goal was to evaluate the status of Whorf's Linguistic Relativity Hypothesis, the claim that language one speaks determines the way that one speaks. I have argued, following Davidson, that the underlying grammatical structure of a language cannot determine general categories of thought in the manner that Whorf supposed. However, I also argued, this time following Aune, that differences in the referential structures may be correlated with differences in "conceptual schemes". But this is a far cry from Whorf's relativism. First of all, Aune's relativism allows that there can be several conceptual schemes represented within a single language. Secondly, on Aune's theory, language does not cause the conceptual schemes to be adopted by people through some unconscious process (as Whorf holds). Rather, on Aune's theory, people with new conceptual schemes cause changes in the referential structures of the languages of their community.

Although Davidson mounts a powerful criticism of Whorf's Linguistic Relativity Hypothesis, his criticism is actually made stronger when supplemented by a theory like Aune's. That is because Davidson, in arguing against extreme relativism, ends up ruling out any significant variation in thought whatsoever. However, any introductory course in sociology or anthropology (or, for that matter, one's everyday interactions with different people) should be enough to convince one that there are differences in the way different groups of people think. These differences are significant enough and systematic enough to be noteworthy, even if they do not add up to the profound differences that Whorf wrote about. What Aune's theory does is show us that these differences are sometimes reflected in language, even if they are not caused by language. In providing an alternate non-Whorfian explanation for the differences that most of us are willing to recognize, Aune makes Davidson's case against extreme relativism easier to accept.

The purpose of this section is to investigate another way in which we can account for variation in thought without invoking Whorfian explanations. Like Aune's theory, the results of this section show us how we can explain the obvious conceptual variation between various groups without postulating the dubious doctrines of incommensurability and unconscious linguistic influences. However, while Aune was still primarily interested in the relation between language and conceptual schemes, this section will consider aspects of conceptual variation that will turn out to have nothing to do with language. The purpose of raising them is to show, once again, that it is possible to explain "obvious" conceptual variation without resorting to Whorfian extremism.

So, while it would be possible to conclude this work at this point, I want to consider one more "dimension" of conceptual variation and show that while this dimension can be explained, the explanation will have nothing to do with language. This will strengthen the case against Whorf's Linguistic Relativity Hypothesis. However, it should be noted that the issues discussed in this section are treated only very briefly, and that a systematic account of conceptual variation would require a much more detailed examination of these issues.

People often have reasons for their beliefs, their desires and their actions. We give a reason for someone's beliefs, desires or actions by showing that they are a consequence of his set of beliefs and desires. Reasoning, then, can be understood as the following three things:

1. The inference of new beliefs from one's current beliefs.

2. The derivation of new desires from one's current beliefs and desires.

3. The determination of a course of action from one's current beliefs and desires.

In this section I want to consider whether one's reasoning might vary as a function of the language one speaks. Although this version of the Linguistic Relativity Hypothesis was not advanced by Whorf, I want to consider it for two reasons. The first reason is simply that it is an inherently interesting question. The second is that it is not easy to separate out the issue of reasoning from Whorf's version of the Linguistic Relativity Hypothesis in any case. Whorf's claim is that what we think, not how we think, is what is influenced by language. But in section 4.3 we saw that in order to determine what people think, we have to apply a number of constraints when we attempt to derived propositional attitudes from the physical evidence. According to David Lewis, one of those constraints is the Principle of Rationality, which is nothing other than a statement of how we think. (We noted that a theorist like Stephen Stitch would be happier replacing the Principle of Rationality with a "Principle of Causality", but this is just Stitch's alternate statement of "how we think".) In other words, reflection on the problem of radical interpretation shows that the issue of what we think cannot be separated from the issue of how we think.

So the question is, could the language we speak determine how we think? (Where by 'thinking' I mean reasoning, as characterized by the three points above.) This version of the Linguistic Relativity Hypothesis is not clearly addressed by Davidson's critique of conceptual relativism, and therefore deserves special attention.

We begin by noting that reasoning has a normative element. The question 'how ought we think?' is very different from the question 'how do we think?'. If someone thinks exactly the way they ought to we say that that person is "rational". Of course people often are not rational, but it is interesting to note that when we describe such a person's reasoning we use expressions like 'not rational', 'mistaken', 'stupid', 'in error', 'faulty', etc., all of which are normative terms. In other words, when we describe how people do think, we often describe that thinking in terms of how well it conforms to the norm of perfect rationality. In answering the question of how people do think, we often presume an answer to the question of how people ought to think, and use that answer as part of our description of how they do think.

It is clear that people do not always reason in the same way. One person answers '46' to the arithmetic question and the next person answers '48'. People with similar socio-economic and educational backgrounds can have wildly different political views. Some of us are theists and some are not. These are differences in reasoning that we can observe between individuals in our own culture. But we can easily imagine that there are two cultures, and in one the individuals tend to reason one way, and in the other another way. If we found such a cross-cultural difference, could we ascribe the differences to language?

Before rushing into this question, let us stop to consider how we should understand the differences in reasoning that have been postulated between the two groups. The paragraph before last pointed out that the norm of rationality has something to do with how we describe actual reasoning. However, the interplay between normative and descriptive factors is complex, and consequently there are at least four different ways that we can describe differences in reasoning between different people. I shall characterize them as four different "theories" to explain the existence of alternate patterns of reasoning. I shall give these "theories" the following names:

1. The Performance Error Theory,
2. The Theory of Systematic Irrationality,
3. The Theory of Relativity of Rationality, and
4. The Socio-Historical Emergence Theory of Rationality.

The Performance Error Theory

Donald Davidson, David Lewis and Daniel Dennett all hold that rationality is constitutive of an epistemic agent. By this I mean that they hold that any thing has beliefs (an epistemic agent) only if it is rational, and furthermore, it is an epistemic agent only because it is rational. It must be rational because the ascription of irrational belief systems (for example, belief systems with inconsistencies) undermines the very ascription of beliefs. Systematic irrationality is impossible.

Of course all these theorists recognize that people reason differently. So how do they account for the differences? The answer is simple: If two people (or two societies) reason differently, then at least one of them must be making a performance error. I am here invoking Chomsky's distinction between performance and competence. Chomsky holds that all English speakers have perfect competence over their particular idiolect of English, but that each speaker is subject to continual performance errors. Similarly, Davidson, Lewis and Dennett must hold that each epistemic agent has a perfectly rational competence for belief inference, but that each of us is subject to performance errors in greater or lesser degrees.

It is possible, I suppose, to advocate a version of the Linguistic Relativity Hypothesis based on this theory of variance of reasoning, but it would not have the force that Whorf originally intended with his Linguistic Relativity Hypothesis. It would not state that one's competence to reason is dependent on language, rather, it would state that one's performance of particular ratiocinations is sometimes subject to interference by linguistic factors.

The Theory of Systematic Irrationality

Another theory of variance in reasoning is that two people differ in their reasoning, one or both of them may be systematically irrational. That is, their very "competence" to reason is characterized by a systematic irrationality. Of course, for Davidson, Lewis and Dennett this is out of the question, for they hold that if someone were systematically irrational that would undercut our very ability to ascribe beliefs to them, and therefore they could not be construed as reasoning in any fashion, rational or irrational.

However, not all theorists agree that rationality is constitutive of an epistemic agent in this way. Stephen Stitch, for example, has argued that the ascription of beliefs has nothing to do with rationality, rather, he holds that we assign beliefs according to their causal role.[382] Since, according to Stitch, our method of ascribing beliefs is independent of considerations of rationality, the way is open for us to conclude that people are systematically irrational, provided that (1) we have an independent standard of rationality, and (2) empirical evidence shows that people do systematically depart from that standard. Stitch argues that both of these conditions hold. Regarding the second point, Stitch believes that much of the psychological research in recent decades provides clear empirical evidence that irrationality is rampant.

I believe that this has been a very popular approach to understanding variations in patterns of reasoning. The basic idea is that some of us reason the right way some of the time.

---

[382] Actually Stitch's theory is considerably more complex than this, but we need not concern ourselves with the details here. See S. Stitch, From Folk Psychology to Cognitive Science, (The M.I.T. Press, 1983).

but many of us reason the wrong way, perhaps much of the time. These are not merely performance errors. Even when we make no mistakes according to our own lights we are still wrong, because our way is wrong. I think that this view has been held by a broad range of theorists, including Roger Bacon, Levy-Bruhl, Evans-Pritchard, Pareto, Freud, Piaget and very interesting recent versions by Gilbert Harman and Christopher Cherniak.[383] According to these last two theorists, all human thought is systematically sub-rational when measured against a standard of ideal rationality.

Can this approach by linked to the Linguistic Relativity Hypothesis? Indeed, it has been, by Roger Bacon:

> Four species of idols beset the human mind, to which (for distinction's sake) we have assigned names, calling the first Idols of the Tribe, the second Idols of the Den, the third Idols of the Market, the fourth Idols of the Theater.
>
> The formation of notions and axioms on the foundation of true induction is the only fitting remedy by which we can ward off and expel these idols. It is, however, of great service to point them out, for the doctrine of idols bears the same relation to the interpretation of nature as the confutation of sophisms does to common logic....
>
> There are...idols formed by the reciprocal intercourse and society of man with man, which we call Idols of the Market, from the commerce and association of men with each other; for men converse by means of language, but words are formed at the will of the generality, and there arises from a bad and unapt formation of words a wonderful obstruction to the mind. Nor can the definitions and explanations with which learned men are wont to guard and protect themselves in some instances form a complete remedy - words still manifestly force the understanding, throw everything

---

[383] G. Harman, Change in View, (The M.I.T. Press, 1986), and C. Cherniak, Minimal Rationality, (The M.I.T. Press, 1986).

> into confusion and lead mankind into vain and innumerable controversies and fallacies.[384]

So a version of the Linguistic Relativity Hypothesis is possible given this approach to variation in reasoning, but note that the effect of language on reasoning is entirely negative on this account.

The Theory of Relativity of Rationality

Both of the previous approaches assumed that there is one normative standard of rationality for all persons and societies. A third approach holds that normative standards of rationality are relative to the social practises of a particular culture. That is, the normative basis of rationality is a "natural" phenomenon, and theories that appeal to a transcendent version of rationality have just got it wrong. Perhaps the most well known recent advocate of this approach is Peter Winch who adapted Wittgenstein's extreme nominalism to the concept of rationality.[385] Other advocates of this approach include Neitzsche, Michel Foucault, Thomas Kuhn, and a number of contemporary sociologists of knowledge, including Barnes and Bloor.

Assume two societies appear to reason in different ways regarding some issue. With the previous two approaches we had to explain this difference by postulating that one or both societies is wrong in their reasoning, either because of performance errors or their very way of reasoning is flawed. But given Winch's relativistic approach we can come up with a radically different explanation. We can say that both societies have reasoned correctly, each according to its own standards of rationality. Of course, individual thinkers may still violate the norms of rationality within their own community. But on this approach it is not possible that a whole community could be wrong in the way that they reason.

---

[384] F. Bacon, Novum Organum, (Colonial Press, 1900), excerpts in J. E. Curtis and J. W. Petras, eds., The Sociology of Knowledge, (Praeger Publishers, 1970), 89, 90.

[385] P. Winch, "Understanding a Primitive Society", American Philosophical Quarterly, 1964, 1, 307-324.

How does this approach to variations in reasoning fit with the Linguistic Relativity Hypothesis? It allows for a very powerful version of linguistic relativity, for we can now hypothesize that language has a lot to do with the standard of rationality that is used by a particular speech-community. That is, this version of the Linguistic Relativity Hypothesis will state that the very norms of rationality within a speech-community are determined by the language spoken by that speech-community.

The Socio-Historical Emergence Theory of Rationality

There is a middle ground position between Winch's relativistic approach to rationality, and the monistic view of rationality endorsed by the first two positions. This middle ground position is held by Peirce, Ian Hacking, Alisdair MacIntyre, and Hilary Putnam. I will focus on Hacking's formulation in what follows.

The middle ground position is that standards of rationality are socio-historical products; human creations that have definite starting points and histories. Hacking, following the historian A. C. Crombie, prefers to call these developments "styles of reasoning". The following passage from Crombie should give some idea of what is meant by a "style of reasoning":

> The active promotion and diversification of the scientific methods of late medieval and early modern Europe reflected the general growth of a research mentality in European society, a mentality conditioned and increasingly committed by its circumstances to expect and to look actively for problems to formulate and solve, rather than for an accepted consensus without argument. The varieties of scientific methods so brought into play may be distinguished as,
> (1) the simple postulation established in the mathematical sciences,
> (2) the experimental exploration and measurement of more complex observable relations,
> (3) the hypothetical construction of analogical models,
> (4) the ordering of variety by comparison and taxonomy,
> (5) the statistical analysis of regularities of populations and the calculus of probabilities, and
> (6) the historical derivation of genetic development.
>
> The first three of these methods concerns essentially the science of individual regularities, and the second three the science of the regularities of populations ordered in time and space.[386]

Hacking's view is that each style of reasoning establishes its own standards of rationality, and furthermore, the style introduces a whole new class of statements as candidates for truth-or-falsehood. However, this is not to be confused with the total relativism of Winch, as the following passage from Hacking should make clear:

---

[386] A. C. Crombie, "Philosophical Presuppositions and Shifting Interpretations of Galileo", in J. Hintikka, D. Gruender, and E. Agazzi, Theory Change, Ancient Axiomatics and Galileo's Methodology, (D. Reidel, 1981), 284.

> Consider Hamlet's maxim, that nothing's either good or bad but thinking makes it so. If we transfer this to truth and falsehood, this is ambiguous between (a) Nothing, which is true, is true, and nothing, which is false, is false, but thinking makes it so: (b) Nothing's either true-or-false but thinking makes it so. It is (b) that preoccupies me. My [view] is...that the sense of a proposition p, the way in which it points to truth or falsehood hinges on the style of reasoning appropriate to p...
>
> The distinction between (a) and (b) furnishes a distinction between subjectivity [Hacking's term for the extreme relativism of someone like Winch] and relativity. Let (a) be subjectivism: by thinking we might make something true or false. Let (b) be the kind of relativity that I address in this paper: by thinking [more precisely, by introducing a new style of thought] new candidates for truth and falsehood may be brought into being... For my part, I have no doubt that our discoveries are 'objective', simply because the styles of reasoning that we employ determine what counts as objectivity...
>
> Can there not be a meta-reason justifying a style of reason? Can one not, for example, appeal to success? It need not be success in generating technology, although that does matter. Nor is it to be success in getting at the truth, for that would be circular. There can, however, be non-circular successes in truth-related matters. For example, following Imre Lakatos, one might revamp Popper's method of conjecture and refutation, urging that a methodology of research programmes constantly opens up new things to think about. I have quoted Chomsky giving a similar meta-reason. On his analysis of the Galilean style, it has not only worked remarkably well, but also, in the natural sciences at least, we have no alternative but to go on using that style, although, of course, in the future it may not work. Although Chomsky does not make the distinction, his meta-reason is less that Galileo's style continues to find out the truth about the universe than that it poses new kinds of probing and answering. It has produced an open-ended dialogue. That might terminate in the face of a nature that ceased to participate in ways that the Galilean can make sense of. We know it might cease to cater to our interests, but at present (says Chomsky) we have no alternative.[387]

---

[387] I. Hacking, "Language, Truth and Reason", in M. Hollis and S. Lukes, eds., Rationality and Relativism, (The M.I.T. Press, 1982), 49, 65-66.

So the picture that Hacking paints is that throughout history new styles of reasoning appear, bringing in new candidates for truth-or-falsity. Not all styles of reasoning are equal in stature though, for there is a meta-criterion, "fecundity", that establishes some styles as superior to others.

One other point about Hacking's approach to reasoning: He does not claim that standards for deduction and induction are relative to a style of reasoning. Deduction and (even) induction are, according to Hacking, simply rules for jumping from truth to truth within a closed class of statements. To add new statements, that is, to add new candidates for truth-and-falsity, requires the introduction of a new style of thought. So we might say that insofar as deduction and induction are common to all styles of thought they comprise a common core of rationality that runs through all styles of thought. Hacking, however, would be quick to point out that it is a mistake to focus too heavily on truth-preserving calculi.

I do not wish to debate the details of Hacking's proposal, but I want to point out the pattern of his approach. He is stating that norms of rationality emerge in a natural setting, but that extreme relativism is not thereby entailed, because the competing norms are to be measured against some sort of pragmatic meta-criterion. His position is, then, something of a middle ground between the universal standards of Davidson/Lewis/Dennett and Bacon/Pareto/Stitch, and the total relativism of Winch

Is Hacking's approach consistent with the Linguistic Relativity Hypothesis? Superficially it seems that it could be.

for we can hypothesize that each language might be conducive to the formation of particular styles of reasoning, but not others. This is a relatively powerful version of the linguistic relativity hypothesis.

We have examined four different ways of explaining the existence of variation in reasoning between different individuals and societies. Each of them is compatible with a version of the Linguistic Relativity Hypothesis, but the first two approaches support only a very weak (and in my view, somewhat implausible) version in which language never helps, but can only hinder, good reasoning. Only the last two versions are consistent with a really robust version of the Linguistic Relativity Hypothesis. The task at hand, then, is to determine which of these four approaches to rationality is correct.

Are Standards of Rationality A Priori?

If standards of rationality are **a priori**, then either theory 1 or theory 2 must be correct. If standards of rationality are empirical, and we assume that not everyone has the same view of rationality and that further conceptual progress is possible, then either theory 3 or theory 4 is correct. (The second conjunct is required because even if standards of rationality are empirical, it is possible that the first person who ever thought about rationality came up with the optimal theory of rational standards, and every one who thought about it since has concurred with this theory. It is an empirical fact that not everyone agrees in their theories of rational standards, therefore the second conjunct is true.)

An example of a standard of rationality that someone might propose is the following: "An agent should choose the action that will maximize his expected utility." Does this normative statement correctly point us in the direction of rationality? If someone follows this advice, is it true that he is being rational? How can we go about answering these questions?

By 'a priori' I mean a statement that is true and knowable by some method other than empirical investigation. Several philosophical traditions have claimed that there is a priori knowledge, and that there are legitimate techniques for accessing this knowledge. The Rationalist tradition, exemplified by Descartes, Leibniz and Spinoza, attempted to specify precise methodologies for accessing a priori truths, and indeed, their philosophies are essentially specifications of grand a priori truths. Kant offered a somewhat different approach to the a priori, and so did the Phenomenologist school. However, as should be apparent, there has been a rich and continuing historical interest in the a priori.

Empiricism is another tradition that is superficially hostile to the concept of a priori truths. However, it has been repeatedly pointed out that empiricism is plagued by implicit assumptions that essentially play the role of a priori principles.[388]

---

[388] Perhaps nowhere better than in Bruce Aune, Rationalism, Empiricism and Pragmatism, (Random House, 1970).

In the twentieth century a number of philosophers have proposed the so-called "linguistic theory of the a priori", in which a priori truths are held to be analytic statements, that is, statements that are true in terms of meaning alone.

I do not wish to rehearse the many arguments that have been made against the notion of a priori knowledge, and against the linguistic theory of the a priori. I will simply assume that the concept of the a priori is wrong.

However, in rejecting the a priori we must not forget the essential epistemological role that a priori knowledge was held to perform. A priori principles served to provide a structure to knowledge that is simply not there if one attends to the purely empirical. Kant held that causality was an a priori principle, and that it provides an organizing framework from which we can now see various empirical events as causal sequences. There is something correct about this approach, namely, the recognition that any epistemology has to account for our knowledge of the world by separating out two elements: the contribution of the world in terms of its causal interaction with the organism (mainly through the senses), and the contribution of the knowing organism itself. Where Kant and the other a priorists went wrong is in their conclusion that the organism's contribution is fixed and immutable and in some sense transcendental.

A much more acceptable account of the organism's contribution was offered by the pragmatist philosophers, starting with Peirce, and reaching full expression in Quine. On Quine's view, the knowledge we have is profoundly underdetermined by the sensory evidence we receive. What completes our knowledge and gives it its form is not a set of a priori principles, but rather, the scientific heritage and general cultural background that each of us acquires as we are socialized. This heritage is full of arbitrary hypotheses (arbitrary in the sense that others could accommodate the sensory information just as well). Some of the most general of these hypotheses are what Kant mistook for a priori principles. However, if our intellectual history had taken a different turn we might have had different high-level organizing hypotheses. Furthermore, the high-level hypotheses that we now use to give shape to our experience are not immutable. We may collectively decide to modify them over time, although Quine warns that we cannot change them too radically. Using Neurath's metaphor, Quine says that we are in the position of rebuilding our ship at sea, plank by plank.[389] Too drastic a change would spell disaster.

The pragmatic theory of the organism's contribution to his knowledge of the world is far more plausible than the a priori theory. (I am simply stating this conclusion, not arguing for it.) Let us now apply this theory to the issue of standards of rationality. Imagine that there is a community that endorses the claim that an organism is rational only if it chooses actions that will maximize its expected utility. Rather than treating it as an immutable a priori truth, we (as students of that community's culture) can now treat it as a high-level hypothesis that has somehow crept into their collective thinking. If they choose not to look at it critically, it will serve as an organizing principle, allowing them to arrange lower-level observations into a pattern. However, they may, at any time, replace it with another high-level hypothesis, provided that the alternate serves at least as well as organizing the lower-level observations. The new principle may turn out to be a hit, replacing the old. If this happens we will conclude that the standards of rationality have changed over time for that community.

---

[389] W. V. O. Quine, Word and Object, (The M.I.T. Press, 1960), 3.

I think that this is exactly how we should view standards of rationality. They are historical products, and they could have been otherwise. This means that of the four theories of variance in reasoning that we considered, the first two have to be rejected.

Putnam on Why Reason Can't be Naturalized

We are left with two theories, or rather, two types of theories. One type says that rationality is entirely relative to a culture. Winch says that it is incorrect to try to assess the standards of rationality extant in Hopi culture in terms of our standards, and there is no vantage point outside of particular cultures from which the two can be measured. The other type of theory agrees that standards of rationality are nothing more than products of a particular culture, but that there are meta-criteria against which different standards of rationality can be evaluated, and as a result of this evaluation we can conclude that some standards are superior to others. Hacking's theory belongs to this group.

Putnam says of the first type of theory that it is an attempt to completely "naturalize" reason. He has presented a powerful argument against the naturalization of rationality,[390] which goes as follows:

Assume that the norms of rationality are entirely determined by local cultural factors. That means that if Karl is a Hopi, then Karl's belief formation, desire formation and action formation are to be judged as rational or irrational against the norms of Hopi culture. Assume that we are students of Karl's beliefs, desires and actions. That means that we will form a theory of Karl's beliefs/desires/actions as judged against the norms of Hopi culture as judged against the norms of our culture. (This sentence is to be parsed as follows: (That means that we will form a theory (of Karl's beliefs/desires/actions as judged against the norms of Hopi culture) as judged against the norms of our culture)). In other words, our cultural determinist stance has led us to view Hopi culture as a kind of a logical construction from within our culture.

Now assume also that Karl is a student of our beliefs, desires and actions. What are we to make of that? Well, as relativist, we can hold that Karl is going to form a theory of our beliefs/desires/actions as judged against the norms of rationality in our culture as judged against the norms of rationality of Hopi culture. But that means that we are now holding a theory whereby our culture is held to be a logical construction out of Hopi culture, and Hopi culture is now the "absolute reference point" against which all other cultures are measured. This is an untenable position, according to Putnam. He says that it is exactly analogous to the untenable position held by positivist philosophers who held that the world is a logical construction out of sense data.

---

[390] H. Putnam, "Why Reason Can't Be Naturalized", Synthese, 1982, 52, 1-23. Reprinted in K. Baynes, J. Bohman, and T. McCarthy, eds. After Philosophy, (The M.I.T. Press, 1987).

> The 'methodological solipsist' - one thinks of Carnap's Logische Aufbau or of Mach's Analyse der Empfindungen - holds that all our talk can be reduced to talk about experiences and logical constructions out of experiences. More precisely, he holds that everything he can conceive of is identical (in the ultimate logical analyses of his language) with one or another complex of his own experiences. What makes him a *methodological* solipsist as opposed to a real solipsist is that he kindly adds that *you*, dear reader, are the 'I' of this construction when *you* perform it: he says that *everyone* is a (methodological) solipsist.
>
> The trouble, which should be obvious, is that his two stances are ludicrously incompatible. His solipsist stance implies an enormous asymmetry between persons: my body is a construction out of my experiences, in the system, but *your* body isn't a construction out of *your* experiences. It's a construction out of *my* experiences. And your experiences - viewed from within the system - are a construction out of your bodily behavior, which, as just said, is a construction out of *my* experiences. My experiences are different from everyone else's (within the system) in that they are what *everything* is constructed from. But his transcendental stance is that it's all symmetrical: the 'you' he addresses in his higher-order remark cannot be the empirical 'you' of the system. But if it's really true that the 'you' of the system is the only 'you' he can understand, then the transcendental remark is unintelligible. Moral: don't be a methodological solipsist unless you are a *real* solipsist![391]

---

[391] *Ibid.*, 230-231.

Putnam states that there is an exact parallel between the doctrine of methodological solipsism (obviously, this term is being used differently from the recent usage inspired by Jerry Fodor) and the doctrine that rationality is culturally determined. The cultural determinist must hold, as we have seen, that Hopi culture is a logical construction from the point of view of our culture. This establishes an "enormous asymmetry" But the cultural determinist also makes a "higher-order, transcendental" claim that all cultures are in a sense equal, since the other culture can view our culture as a logical construction from their point of view. The problem, once again, is that if you take the first claim that implies asymmetry seriously, then the transcendental claim becomes unintelligible.

Putnam argument against methodological solipsism says, in effect, that methodological solipsism collapses into real solipsism. The parallel argument against the theory that rationality is culturally relative is that that theory collapses into what Putnam calls "cultural imperialism", which is the position that my culture's theory of rationality is the theory of rationality. But this position, although still culturally determinist, is no longer relativist, for it postulates two classes of cultures. First, there is the class of real, living, breathing cultures, which has exactly one member - viz., my culture. It is this culture that determines what rationality is. Then there is the other class, which consist of all other cultures. However, these are not real, living, breathing cultures, but rather, they are logical constructions from the point of view of my culture. Putnam presents an argument that this position is self-refuting,[392] however, I don't think that for our purposes it is necessary to review that argument. It is sufficient to point out that cultural imperialism is no longer relativistic, and it is relativism that we are interested in.

We began by considering two types of theories. One type says that rationality is entirely relative to a culture. The other type of theory agrees that standards of rationality are nothing more than products of a particular culture, but that there are meta-criteria against which different standards of rationality can be evaluated, and as a result of this evaluation we can conclude that some standards are superior to others. On the former type of theory rationality is completely "naturalized", but in the latter it is not. Putnam's argument shows us that the former type of theory is not viable.

Rationality and Language

Let us review where we are. Our original question was: can language affect the way people reason, where reasoning is the inference of new beliefs, desires or actions from one's current stock of beliefs and desires? It was noted that people do vary in their reasoning; we do not all come to the same conclusions from the same premises. There is, then, a phenomenon that we can call variation in reasoning. Four theories of variation in reasoning were considered. The first two views assume that there is one standard of reasoning that will apply to all cultures.

---

[392] Ibid., 233.

One way of defending either of these two claims is to argue that standards of rationality are a priori. However, the idea of a priori knowledge was rejected as unacceptable. The rejection of the a priori led us to the pragmatic theory of rationality. The pragmatic theory states that standards of rationality are a cultural product, and they can change over time. This forces us to restrict our attention to the two remaining theories of variation in reasoning, the relativistic theory and the historically emergent theory of rationality. The former theory is naturalistic, in that it states that cultural norms completely exhaust the content of standards of rationality. The latter theory states that culturally relative standards of rationality are themselves subject to evaluation in terms of meta-criteria. Putnam's argument against the former theory was reviewed, which leaves us with the latter theory.

It might be objected that the theory of historically emergent rationality cannot be distinguished from the relativistic theory of rationality, for what supposedly distinguishes the two, viz., meta-criteria, must either be a priori or a cultural product. Since nothing is a priori, they must be a cultural product, and therefore the theory of historically emergent rationality collapses into the relativistic theory. The premises of this argument are all correct, but the conclusion does not follow. It follows only given the additional premise: Meta-criteria vary between cultures. But this can be denied. We can hold that standards of rationality vary between cultures, but meta-criteria for judging standards of rationality are universal, even though they are cultural products. I will not develop a detailed argument for this, but it is easy enough to see how one could be developed. It would be a Davidsonian/Kantian style argument with Putnamian themes: if meta-criteria of rationality were not universal, then we would be faced with either of two situations, the self-contradictory stance of the cultural relativist or the unacceptably asymmetric stance of the cultural imperialism. The latter position is unacceptable for it does not allow us to conceive of other cultures as having a status equal to our own. Question (in the Davidsonian/Kantian style): What makes it possible to conceive of other cultures as having a status equal to our own? Answer: Universal meta-criteria that can be used to judge local standards of rationality.

Although meta-criteria do not vary, standards of rationality can. Can this variation be due to language? I think not. But first I will point out that if rationality is a function of language, then the version of the Linguistic Relativity Hypothesis that expresses this relation cannot be a scientific claim, as Whorf intended the Linguistic Relativity Hypothesis to be. Science is a style of reasoning that places a high premium on precision, detail and predictability. A scientific version of the claim that rationality varies as language does would have to produce testable hypotheses of the form 'If the members of speech community S spoke the language L, then they would have standards of rationality R'. However, if we could form such hypotheses precisely and in detail, then our hypotheses would contain complete descriptions of standards of rationality that have not yet been used by any speech community. Now, if Hacking is anywhere close to correct, the development of a standard of rationality (or style of reasoning) can be a major impetus to scientific progress. That means that the version of the Linguistic Relativity Hypothesis we are now considering would require us to specify, at least in part, scientific advances before they happen. But as Karl Popper has pointed out, the idea that science can predict future progress in science is incoherent.[393] Consequently, any version of the Linguistic Relativity Hypothesis that claims rationality is a function of language must necessarily be a rather general "philosophical" principle, rather than the detailed, precise scientific claim that Whorf was interested in.

However, I do not believe that even a general "philosophical" version of the claim that rationality is a function of language is warranted. As preparation for this claim, I want to present briefly the recent views of five authors who have discussed the issue of how alternate styles of reasoning, or alternate standards of rationality are introduced.

Hacking on New Styles of Reasoning

Hacking has chosen the word 'style' with care. Like styles in clothing or in music, he holds that styles of reasoning are conscious creations of individuals or groups of individuals. Thales, Galileo and Marx are examples of individuals who have promoted new styles of reasoning.[394] For something to count as a new style of reasoning, it has to be recognized as something new by the members of the community in which it is introduced. It has been put on the table, so to speak, and the members of the community will either adopt it or ignore it. In other words, it is consciously introduced by its advocate, and consciously accepted or rejected by its audience. This gives styles of reasoning a definite historical nature.

---

[393] K. R. Popper, The Poverty of Historicism, (Harper Torchbooks, 1957).

> "There have been different styles of reasoning. Many of these are discernible within our own history. They emerge at definite points and have distinct trajectories of maturation. Some die out, and others are still going strong."[395]

MacIntyre on Reason over Might

Alisdair MacIntyre asks us to consider what might happen if two communities with different standards of rationality happen to come into conflict over some issue.[396] The Nietzschian answer (represented currently by Foucault), which is also a relativistic answer, is that might will make right. There is no higher court of rationality that will settle the dispute, so if the dispute is settled at all it will be settled by power (even if it is power masking as reason).

MacIntyre agrees with the relativist that there is no higher court of rationality, but there can still be a way of arbitrating the dispute without descending to a war of wills. Arbitration is possible if at least one of the parties has an internal standard of rationality such that it is self-critical, and willing to accept that the intellectual history of the culture is such that certain key problems and contradictions have emerged that "cannot be resolved within the particular tradition's own conceptual framework".[397] If the competing culture offers a way around these problems and contradictions, the first culture may choose to embrace the competitor's standard of rationality.

---

[394] I. Hacking, op. cit., 50-51.

[395] Ibid., 64.

[396] A. MacIntyre, "Relativism, Power, and Philosophy", Proceedings and Addresses of the American Philosophical Association, 1985, 5-22. Reprinted in K. Baynes, J. Bohman, and T. McCarthy, eds. After Philosophy, (The M.I.T. Press, 1987).

> What rendered Newtonian physics rationally superior to its Galilean and Aristotelean predecessors and to its Cartesian rivals was that it was able to transcend their limitations by solving problems in areas in which those predecessors and rivals could by their own standards of scientific progress make no progress. So we cannot say wherein the rational superiority of Newtonian physics consisted except historically in terms of its relationship to those predecessors and rivals whom it challenged and displaced. Abstract Newtonian physics from its context, and then ask wherein the rational superiority of one to the other consists, and you will be met with insoluble incommensurability problems. Thus knowing how Newton and the Newtonians actually came to adopt and defend their views is essential to knowing why Newtonian physics is to be accounted rationally superior.[398]

Whereas Hacking suggested that a culture can internally change its standard of rationality, MacIntyre is suggesting that one culture can win another over. But note that both Hacking and MacIntyre see these changes as conscious acts on the part of the members of the respective cultures. These changes are datable (and debatables) historical events. Indeed, MacIntyre's last point in the quotation above is that changes in standards of rationality cannot even be understood unless the historical context is taken into account.

---

[397] Ibid., 407.

[398] Ibid., 416.

Cohen on Conversion in Scientific Revolutions

The scientific historian I. Bernard Cohen has recently published a wide-ranging study of scientific revolutions. Since Kuhn's The Structure of Scientific Revolutions, it has been popular to hold that the members of the pre-revolution and post-revolution schools do not talk to one another. Cohen disagrees with this:

> ...[T]here is one variety of experience in scientific revolutions that occurs again and again in the primary and secondary literature, and I would like to give it some consideration here. That phenomenon is conversion. Max Planck...is often quoted to the effect that 'new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it.' A similar sentiment was expressed a half-century earlier by Harvard's Professor Joseph Lovering, when he told his students that there are two theories of light, the wave theory and the corpuscular. Today, he is said to have remarked, everyone believes in the wave theory; the reason is that all those who believed in the corpuscular theory are dead. There is a measure of truth in such statements, as we all know, and yet a new scientific idea does win adherents, and even convinces some opponents, as has been seen in many examples throughout this book. Planck himself actually witnessed the acceptance, modification and application of his fundamental concept by his fellow scientists. This feature of scientific revolutions - the winning over of working scientists - is common enough that I have used its magnitude as a mark of the transition from a revolution on paper to a revolution in science.[399]

---

[399] I. B. Cohen, Revolution in Science, (Harvard University Press, 1985), 467-468.

Cohen's work stands as a corrective to the Kuhnian view of scientists talking past one another. Like Hacking and MacIntyre, Cohen sees intellectual progress as the result of a dialogue.

Vickers on Occult and Scientific Mentalities

The historian Brian Vickers recently edited a collection of essays on the interaction of the occult and scientific traditions in the 16th and 17th centuries.[400] These essays provide abundant historical evidence that these two traditions were involved in a rich and complex debate for over two centuries. Often individuals (notably Kepler and Newton) struggled with the appeal of both traditions. Some of the contributors, including Vickers himself, hold that there are deep and fundamental distinctions between the two styles of thought, to the extent that the occultists could not really understand the claims of the scientists. Other contributors disagree, holding that the supposed differences have often been overemphasized. But both groups point out that the dialogue was rich, and there were many cases of individual conversions. The volume supports the picture of intellectual change that we find in Hacking, MacIntyre and Cohen.

Darnton on Folk Tales

Historian Robert Darnton's The Great Cat Massacre is a study of modes of thought in 18th century France, from that of the peasant to the artisan to the bourgeois to the philosopher. Several of the chapters of the book contribute to the theme I am trying to develop, but I will just consider one: the chapter on folk tales in peasant life. Darnton writes:

---

[400] B. Vickers, ed., Occult and Scientific Mentalities in the Renaissance, (Cambridge University Press, 1984).

> The peasants of the Old Regime...tried to make sense of the world, in all its blooming, buzzing confusion, with the materials they had at hand. Those materials included a vast repertory of stories derived from ancient Indo-European lore. The peasant tellers of tales did not merely find the stories amusing or frightening or functional. They found them 'good to think with.' They worked with them in their own manner, using them to piece together a picture of reality and to show what that reality meant for people at the bottom of the social order. In the process they infused the tales with many meanings, most of which are now lost because they were embedded in contexts and performances that cannot be recaptured. At a general level, however, some of the significance still shows through the texts. By studying the entire corpus of them and by comparing them with corresponding tales in other traditions, one can see this general dimension of meaning expressed in characteristic narrative devices - ways of framing stories, setting tone, combining motifs, and inflecting plots. The French tales have a common style, which communicates a common way of construing experience. Unlike the tales of Perrault, they do not provide morals; and unlike the philosophies of the Enlightenment, they do not deal in abstractions. But they show how the world is made and how one can cope with it. The world is made of fools and knaves, they say: better to be a knave than a fool.[401]

According to Darnton, then, telling folk tales is a style of thought. French peasants took the tools that were available to them, and through a definite historical process, used those tools to accomplish their own intellectual ends, and thereby created their style of thought. Like the other authors mentioned above, Darnton stresses dialogue and intentional human creativity in the creation of a new style.

---

[401] R. Darnton, The Great Cat Massacre, (1984, reprint Random House, 1985), 64.

Whorf felt that language works its effects in a subtle unnoticed way. If Whorf had specifically addressed styles of thought or standards of reasoning, he would have claimed that language determines these things without the members of the speech community noticing that they are so effected. However, the five authors briefly reviewed present a completely incompatible picture of how new styles of thought (or standards of reasoning) develop. These authors stress that a new style of thought emerges through a dialogue, often a heated one with champions and opponents. Far from unnoticed, the emergence of a new style of thought requires conscious decisions from the members of the speech community. And rather than linguistic changes causing changes in styles of thought, these authors suggest that the proponent of a style of thought will change language in order to advance his case. Even the peasants of 18th century France are pictured more as the masters of language than the other way around.

Assembling opposing points of view does not represent an argument, of course. I have no decisive argument against the version of the Linguistic Relativity Hypothesis that states that reasoning or styles of thought in a speech community are functions of the language spoken in that community. However, this hypothesis, as mentioned above, seems destined to remain a rather general "philosophical" doctrine, without definite implications. On the other hand, the view that people consciously create new styles of thought through dialogue and the manipulation of the linguistic materials available to them is a thesis that can be supported through detailed historical studies. At the very least, we can safely conclude that at the present time the thesis that people cause new styles of thought has a lot more going for it (in terms of theoretical interest and empirical results) than the thesis that languages cause styles of thought. Although I cannot definitively refute the latter thesis, I can see no good reason why anyone would want to support it against its competitor.

4.7 CONCLUSION

In Chapter One Whorf's Linguistic Relativity Hypothesis was interpreted as consisting of the following five ideas:

W1. To think is to employ concepts.

W2. Languages have conceptual structures.

W3. Languages differ.

W4. Language determines thought.

W5. Cross-cultural understanding is difficult/impossible.

In Chapter Two the concept of a natural language was explored. The main issue in Chapter Two was could we come up with a defensible concept of a natural language that was consistent with point W2 above. That is, could we come up with a concept of a natural language that associates a conceptual structure with each natural language. First Bloomfield's views were examined, for Bloomfield was still extremely influential at the time that Whorf wrote, but Bloomfield argued, in opposition to Whorf, that all questions about semantics (and thereby conceptual structures) should be put on hold. It was argued that Bloomfield was wrong, and that semantic considerations have to be taken into account if we want to advance in other areas of linguistic study, for example, syntax. However, Whorf himself offered little in the way of a semantic program. His scattered remarks do seem consistent with the theories of meaning developed by linguists in the sixties, however, so these were examined as an approach to natural language that would be consistent with Whorf's claim. These theories were found wanting, however, and finally truth-conditional semantics was defended as the proper approach for semantics. Therefore Whorf's W2 will have to be reinterpreted in the context of truth-conditional semantics if we are to make any sense of it at all.

Chapter Three examined thought, and in particular, Whorf's W1. It was argued that we can make sense of the "concept" concept only in the context of a propositional attitude theory of thought. However, it was also argued that while a propositional attitude theory of thought is appropriate for sociological theory, it is inappropriate for a psychological theory. If individual psychology is to make any progress it must abandon propositional attitudes and other intentional idioms and adopt the concepts of biology and chemistry. Sociology (including the study of the individual qua societal member), on the other hand, must retain the intentional idioms. Consequently, the Linguistic Relativity Hypothesis must be considered a hypothesis within sociological theory (in this sense) rather than psychological theory.

The arguments of Chapters Two and Three lead us to the following picture of human societies: They are composed of persons, where persons are understood intentionally, that is, in terms of their beliefs, desires and actions. Some actions are communicative actions, and some of those are linguistic acts. A linguistic act (or "speech act") is an act where one party intends to affect the propositional attitudes or actions of another party by means of an utterance, uttered with certain complex intentions. Part of the intention of the utterer is that the audience will react, in part, to the literal meaning of the utterance, where the literal meaning is given by the truth conditions of what is uttered, as given by a formal truth-conditional semantic theory.

The approach to sociological theorizing outlined in the previous paragraph is, of course, intended to be applied to particular societies. Applying the theory requires, as a start, an identification of the beliefs and desires of the members of the society, as well as the truth-conditions of the sentences of the language they use. Obviously, if the Linguistic Relativity Hypothesis is to be verified, this belief-meaning identification must take place.

Chapter Four begins by considering the task of belief-meaning identification. It was concluded that belief and meaning have to be identified in tandem. Also, Davidson's views on belief-meaning identification, or "interpretation", to use the standard term, were considered. Davidson attempted to show that radical incommensurability of belief-meaning between different societies is impossible. While Davidson's argument fails, something very similar can be defended, that is, if we can recognize that a group of organisms has beliefs and meanings, then we will always be able to interpret those beliefs and meaning in detail. This shows that any version of the Linguistic Relativity Hypothesis that supports radical incommensurability is wrong.

What about a version of the Linguistic Relativity Hypothesis that is weaker, a version that claims merely that thought varies in significant, although still commensurable, ways as a function of language? We saw that Davidson argues that this too is impossible. In effect he poses a Kantian question: How is interpretation possible? Davidson's transcendental conclusion is that all thinkers, no matter what their linguistic and cultural background is, must be in agreement on most things. If Davidson is correct on this issue then the Linguistic Relativity Hypothesis, even the weakened version we have been considering, is simply false.

However, we saw that Bruce Aune argued that Davidson has too narrow a view of how to compare thoughts of different groups. Rather than looking only at beliefs and sentences - focussing at the level of the proposition, so to speak - Aune argues that we can also focus at a lower level, at the level of objects and properties and at the level of names and predicates. If we focus at this lower level we will be able to detect variation in thought that is invisible to Davidson. However, Aune's suggestion lead us into one of the "Great Debates" of contemporary philosophy of language, the question of the "reality" of this lower level of names and predicates; i.e., the question of the reality of reference. Our conclusion was that with care, we can treat the semantics of reference structures as having the same empirical and theoretical status as the semantics of sentences. This vindicates Aune's approach and allows us to see more contrast between the thoughts of different cultures than Davidson is able to see.

Can Aune's approach support a version of the Linguistic Relativity Hypothesis? It seems not, for on Aune's theory, it is thought that causes language, rather than the other way around. That is, Aune holds that new reference schemes have to be consciously introduced by individuals. These individuals are introducing the schemes precisely because they are intending to propagate a way of describing nature that they already have in mind. Furthermore, in defending Aune's approach we had to bring the causal theory of reference into play. The causal theory of reference, with its "baptisms" of independently identified objects, also supports the view that thought causes language, rather than the other way around.

Finally, our focus shifted to thinking rather than merely inventorying thoughts. Not thinking in a psychological sense, but thinking in the "sociological" sense of reasoning - forming new beliefs, desires, and actions in light of one's present beliefs and desires. Might we not have a version of the

Linguistic Relativity Hypothesis that states that reasoning varies as a function of language?

The discussion began by assuming that reasoning does vary. How might this be accounted for? A number of theories were considered, but only one was supported: a theory that says that standards of rationality are cultural phenomena, but that claims between alternate standards of rationality can be mediated by meta-criteria such as intellectual fecundity. This theory seems, on the face of it, to be compatible with the relativistic claim that reasoning might vary as a function of language. However, by examining the work of a number of authors who advocate this socio-historical emergence theory of rationality we see that they argue that language is a function of new modes of reasoning, rather than the other way around. So although this socio-historical emergence theory supports a notion of intellectual variation that Davidson is blind to, it does not appear to support a version of the Linguistic Relativity Hypothesis.

The case against the Linguistic Relativity Hypothesis can be strengthened by returning to the considerations outlined in the beginning of section 4.1. In section 4.1 I argued that the method of interpretation used to identify beliefs and meanings is critical to the Linguistic Relativity Hypothesis. I pointed out that there are at least four reasons why the method might fail to support the Linguistic Relativity Hypothesis. All four of these possibilities have been shown to be actualities, as I will briefly review.

1. The attribution of language and thought cannot be separated into two distinct methodological undertakings. Rather, meanings and beliefs are jointly attributed in a single methodological undertaking. It is difficult to see how the strong causal relation between language and thought that Whorf suggested can be consistent with the methodological conclusions that we have reached.

2. Our method of interpretation shows that incommensurability, a key feature of Whorf's thesis, is simply not coherent. Davidson's argument against incommensurability stands even in light of the concessions that we have made to variability of referential structures and styles of reasoning.

3. Our method of interpretation, even when supplemented with considerations regarding reference and reasoning, allows for as significant a variation in thought within a linguistic community as between linguistic communities.

4. Finally, our method of interpretation differs from the usual methodological standards in empirical science. For one thing, normative standards play a large role in interpretation. Normative standards of reasoning, for example, are required in order to determine what beliefs a person has. If two interpreters have slightly different norms as to what constitutes the most rational action in a certain situation, then they will attribute different beliefs and desires to subjects who they observe in that situation. Furthermore, as repeatedly pointed out by Davidson, the interpreter can play meanings and beliefs off against one another so that he can come up with alternate

belief-meaning interpretations of the same behavior. These methodological standards are rather "loose" compared to the standards employed in the laboratories of physicists and physiological psychologists. This means that the Linguistic Relativity Hypothesis cannot be viewed as part of hard science, which is, of course, how Whorf thought it should be viewed.

The Linguistic Relativity Hypothesis has not been definitively refuted in these pages. That is not a fault, because true knock-down arguments are extremely rare in any case. Gödel really did knock down Hilbert's formalist philosophy of mathematics, but has anyone really refuted Cartesian Rationalism or proved that God does not exist? My goal has not been definitive refutation or the development of a proof. My goal has been to present arguments that make a case against the Linguistic Relativity Hypothesis, and I hope I have done that.

BIBLIOGRAPHY

Aristotle. *Categories and De Interpretatione.* Translated by J. L. Ackrill. Oxford University Press, 1963.

Aune, B. *Rationalism, Empiricism and Pragmatism.* Random House, 1970.

Aune, B. "Conceptual Relativism." In *Philosophical Perspectives 1: Metaphysics*, edited by J. E. Tomberlin. Ridgeview Publishing, 1987.

Austin, J. L. *How to Do Things With Words.* Harvard University Press, 1962.

Bach, E. *Syntactic Theory.* Holt, Rinehart and Winston, 1974.

Bacon, F. *Novum Organum.* Colonial Press, 1900. Excerpts in *The Sociology of Knowledge*, edited by J. E. Curtis and J. W. Petras. Praeger Publishers, 1970.

Barnes B. and D. Bloor. "Relativism, Rationalism and the Sociology of Knowledge." In *Rationality and Relativism*, edited by M. Hollis and S. Lukes. The M.I.T. Press, 1982.

Baynes, K. J. Bohman, and T. McCarthy, eds. *After Philosophy.* The M.I.T. Press, 1987.

Bennett, J. *Linguistic Behavior.* Cambridge University Press, 1976.

Berlin, B. and Kay, K. *Basic Color Terms: Their Universality and Evolution.* University of California Press, 1969.

Bierwisch, M. "Semantics." In *New Horizons in Linguistics*, edited by J. Lyons. Penguin, 1970.

Birdwhistell, R. L. *Kenesics and Context.* Ballantine Books, 1970.

Bloch, B. "Studies in Colloquial Japanese IV: Phonemics." *Language*, 1950, 26, 86-125. Reprinted in Joos, 1957.

Block, N. ed. Readings in the Philosophy of Psychology. Volume 1. Harvard University Press, 1980.

Block, N. ed . Readings in the Philosophy of Psychology. Volume 2. Harvard University Press, 1981.

Bloomfield, L. "A Set of Postulates for the Science of Language." Language, 1926, 2, 153-164. Reprinted in Joos, 1957.

Bloomfield, L. Language. 1933. Reprint, University of Chicago Press, 1984.

Boaz, F. "Introduction to the Handbook of American Indian Languages." Reprinted in Language, Culture and Society, edited by B. G. Blount. Winthrop Publishers, 1974.

Brentano, F. Psychologie vom Empirischen Standpunkt. 1874.

Brown, R. W., and Lenneberg, E. H. "A Study of Language and Cognition." Journal of Abnormal and Social Psychology, 1954, 49, 454-462.

Brown, R. W. and Lenneberg, E. H. "Studies in Linguistic Relativity." Readings in Social Psychology. 3rd edition, edited by E. E. Maccoby, T. M. Newcomb, and E. L. Hartley. Holt, Rinehart and Winston, 1958.

Bruner, J. S., J. L. Goodnow, and G. A. Austin. A Study of Thinking. John Wiley and Sons, 1956.

Carnap, R. Meaning and Necessity. University of Chicago Press, 1947.

Carroll, J. B. "Introduction." In B. L. Whorf, Language, Thought and Reality, edited by J. B. Carroll. The M.I.T. Press, 1956.

Carroll J. B., and Casagrande, J. B. "The Function of Language Classification in Behavior." Readings in Social Psychology. 3rd edition, edited by E. E. Maccoby, T. M. Newcomb, and E. L. Hartley. Holt, Rinehart, and Winston, 1958.

Cassirer, E. The Philosophy of Symbolic Forms. 3 Volumes. Yale University Press, 1955, 1957.

Cassirer, E. An Essay on Man. Yale University Press, 1962.

Cherniak, C. Minimal Rationality. The M.I.T. Press, 1986.

Chisholm, R. "Intentionality." In The Encyclopedia of Philosophy, edited by P. Edwards. MacMillan Publishing, 1967.

Chomsky, N. "Three Models for the Description of Language." I.R.E. Transactions on Information Theory, 1956, vol. IT-2.

Chomsky, N. "Review of Verbal Behavior by B. F. Skinner." Language, 1959, 35, 26-58.

Chomsky, N. Aspects of the Theory of Syntax. The M.I.T. Press, 1965.

Chomsky, N. "Recent Contributions to the Theory of Innate Ideas." In Boston Studies in the Philosophy of Science, Volume III. The Humanities Press, 1968. Reprinted in The Philosophy of Language, edited by J. R. Searle. Oxford University Press, 1971.

Chomsky, N. Syntactic Structures. Mouton, 1972.

Chomsky, N. Reflections on Language. Pantheon, 1975.

Churchland, P. M. Scientific Realism and the Plasticity of Mind. Cambridge University Press, 1979.

Churchland, P. M. Matter and Consciousness. The M.I.T. Press, 1984.

Cohen, I. B. Revolution in Science. Harvard University Press, 1985.

Cornman, J. W. "Intentionality and Intensionality." Philosophical Quarterly, 1950, 12, 44-52. Reprinted in Intentionality, Mind and Language, edited by A. Marras. University of Illinois Press, 1972.

Crombie, A. C. "Philosophical Presuppositions and Shifting Interpretations of Galileo." In Theory Change, Ancient Axiomatics and Galileo's Methodology, edited by J. Hintikka, D. Gruender, and E. Agazzi. D. Reidel, 1981.

Culicover P. W. and K. Wexler. "Some Syntactic Implications of a Theory of Language Learnability." In Formal Syntax, edited by P. W. Culicover, T. Wasow, and A. Akmajian. Academic Press, 1977.

Culler J. "Introduction." In F. de Saussure, Course in General Linguistics. 1916, reprint Fontana, 1974.

Daly, R. T. Applications of the Mathematical Theory of Linguistics. Mouton, 1974.

Dancy, J. Introduction to Contemporary Epistemology. Basil Blackwell, 1985.

Darnton, R. The Great Cat Massacre. 1984, reprint Random House, 1985.

Davidson, D. "Actions, Reasons, and Causes." Journal of Philosophy, 1963, 60, 685-700. Reprinted in Davidson, 1980.

Davidson, D. "The Logical Form of Action Statements." In The Logic of Decision and Action, edited by N. Rescher. University of Pittsburgh Press, 1967. Reprinted in Davidson, 1980.

Davidson, D. "Truth and Meaning." Synthése, 1967, 17, 304-323. Reprinted in Davidson, 1984.

Davidson, D. "On Saying That." Synthése, 1968, 19, 130-146. Reprinted in Davidson, 1984.

Davidson, D. "Mental Events." In Experience and Theory, edited by L. Foster and J. W. Swanson. University of Massachusetts Press, 1970. Reprinted in D. Davidson, 1980.

Davidson, D. "Agency." In Agent, Action, and Reason, edited by R. Binkley, R. Bronaugh, and A. Marras. University of Toronto Press, 1971. Reprinted in Davidson, 1980.

Davidson, D. "Radical Interpretation." Dialectica, 1973, 27, 313-328. Reprinted in Davidson, 1984.

Davidson, D. "On the Very Idea of a Conceptual Scheme." Proceedings and Addresses of the American Philosophical Association, 1974, 47. Reprinted in Davidson, 1984.

Davidson, D. "Belief and the Basis of Meaning." Synthése, 1974, 27, 309-323. Reprinted in Davidson, 1984.

Davidson, D. "Thought and Talk." In Mind and Language, edited by S. Guttenplan. Oxford University Press, 1975. Reprinted in Davidson, 1984.

Davidson, D. "Reality Without Reference." Dialectica, 1977, 31, 247-253. Reprinted in Davidson, 1984.

Davidson, D. "The Inscrutability of Reference." The Southwestern Journal of Philosophy, 1979, 10, 7-19. Reprinted in Davidson, 1984.

Davidson, D. "Toward a Unified Theory of Meaning and Action." Grazer Philosophische Studien, 1980, 11, 1-12.

Davidson, D. Essays on Actions and Events. Oxford University Press, 1980.

Davidson, D. "Communication and Convention." In D. Davidson, 1984.

Davidson, D. Inquiries into Truth and Interpretation. Oxford University Press, 1984.

Davidson, D. "Empirical Content." In Truth and Interpretation, edited by E. LePore. Basil Blackwell, 1986.

Davidson D. and G. Harman, eds. The Logic of Grammar. Dickenson, 1975.

DeGeorge, R. and F., eds. The Structuralists from Marx to Lévi-Strauss. Anchor Books, 1972.

Dennett, D. C. "A Cure For the Common Code." In Brainstorms, by D. Dennett. The M.I.T. Press, 1978.

Dennett, D. C. Brainstorms. The M.I.T. Press, 1978.

Dennett, D. "Beyond Belief." In Thought and Object, edited by A. Woodfield. The M.I.T. Press, 1982. Reprinted in Dennett, 1987.

Dennett, D. C. "The Language of Thought Reconsidered." In Dennett, 1987.

Dennett, D. C. The Intentional Stance. The M.I.T. Press, 1987.

Derwing, B. Transformational Grammar as a Theory of Language Acquisition. Cambridge University Press, 1973.

Devitt, M. Designation. Columbia University Press, 1981.

Devitt M. and K. Sterelny. Language and Reality. The M.I.T. Press, 1987.

Donnellan, K. "Reasons and Causes." In The Encyclopedia of Philosophy, edited by P. Edwards. MacMillan, 1967.

Dummett, M. Frege: Philosophy of Language. Harper and Row, 1973.

Dummett, M. "The Philosophical Basis of Intuitionistic Logic." In Truth and Other Enigmas, by M. Dummett. Duckworth, 1978.

Eagleton, T. Literary Criticism. Basil Blackwell, 1983.

Eco, U. M. Santambrigio, and P. Violi, eds. Meaning and Mental Representations. Indiana University Press, 1988.

Evans G. and J. McDowell. "Introduction." In Truth and Meaning, edited by G. Evans and J. McDowell. Oxford University Press, 1976.

Evans G. and J. McDowell, eds. Truth and Meaning. Oxford University Press, 1976.

Fauconnier, G. Mental Spaces: Aspects of Meaning Construction in Natural Language. The M.I.T. Press, 1984.

Feyerabend, P. K. "How to Be a Good Empiricist - A Plea for Tolerance in Matters Epistemological." In *Philosophy of Science, The Deleware Seminar*, edited by B. Baumrin. The University of Deleware, 1963. Reprinted in *Challenges to Empiricism*, edited by H. Morick. Wadsworth Publishing, 1972.

Field, H. "Tarski's Theory of Truth." *Journal of Philosophy*, 1972, *69*, 347-375.

Field, H. "Conventionalism and Instrumentalism in Semantics." *Noûs*, 1975, *9*, 375-405.

Fishman, J. A. "A Systemization of the Whorfian Hypothesis." Reprinted in *Culture and Cognition*, edited by J. W. Berry and P. R. Dasen. Metheun, 1974.

Flanagan, O. J. Jr. *The Science of the Mind.* The M.I.T. Press, 1984.

Fodor, J. A. "Could Meaning be an r_m?" *The Journal of Verbal Learning and Verbal Behavior*, 1965, *4*, 73-81.

Fodor, J. A. *Psychological Explanation.* Random House, 1968.

Fodor, J. A. *The Language of Thought.* Thomas Y. Crowell, 1975.

Fodor, J. A. "Computation and Reduction." In *Minnesota Studies in the Philosophy of Science, Volume 9: Perception and Cognition*, edited by W. Savage. University of Minnesota Press, 1978. Reprinted in Fodor, 1981.

Fodor, J. A. "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology." *The Brain and Behavioral Sciences*, 1980, *3*. Reprinted in Fodor, 1981.

Fodor, J. A. *Representations.* The M.I.T. Press, 1981.

Fodor, J. A. *Psychosemantics.* The M.I.T. Press, 1987.

Fodor, J. A., T. G. Bever, and M. F. Garrett. *The Psychology of Language.* McGraw-Hill, 1974.

Fodor J. A. and N. Block. "What Psychological States are Not." *Philosophical Review*, 1972, *81*. Reprinted in Fodor, 1981.

Fodor, J. D. *Semantics: Theories of Meaning in Generative Grammar.* Harvard University Press.

Frege, G. "On Sense and Reference." *Zeitschrift für Philosophie und Philosophische Kritik*, 1892, *100*, 25-50. Original in German, translated in *Philosophical Writings of Gottlob Frege*, edited by P. Geach and M. Black. Basil Blackwell, 1977.

Frege, G. The Foundations of Arithmetic. Trans. by J. L. Austin. Oxford University Press, 1950.

Frege, G. "The Thought: A Logical Inquiry." Tr. by A. M. and M. Quinton, Mind, 1956, 65, 289-311. Reprinted in Philosophical Logic, edited by P. F. Strawson. Oxford University Press, 1967.

Freund, J. The Sociology of Max Weber. 1968, reprint Vintage Books, 1969.

Furth, H. "The Influence of Language on the Development of Concept Formation in Deaf Children." Journal of Abnormal and Social Psychology, 1961, 63, 386-389.

Gazdar, G., et. al. Generalized Phrase Structure Grammar. Harvard University Press, 1985.

Gold, E. M. "Language Identification in the Limit." Information and Control, 1967, 6, 441-474.

Goldman, A. I. A Theory of Human Action. Princeton University Press, 1970.

Goodman N. Contribution to the "Symposium on Innate Ideas." In Boston Studies in the Philosophy of Science, Vol. III. The Humanities Press, 1968.

Goodman, N. Fact, Fiction, and Forecast. 4th ed. Harvard University Press, 1983.

Grice, H. P. "Meaning." Philosophical Review, 1957, 66, 377-388.

Gunderson, K. ed. Language, Mind, and Knowledge: Volume VII, Minnesota Studies in the Philosophy of Science. University of Minnesota Press, 1975.

Hacking, I. "Language, Truth and Reason." In Rationality and Relativism, edited by M. Hollis and S. Lukes. The M.I.T. Press, 1982.

Hamilton, A. G. Logic for Mathematicians. Cambridge University Press, 1978.

Harman, G. "Deep Structure as Logical Form." In Semantics of Natural Language, edited by D. Davidson and G. Harman. D. Reidel, 1972.

Harman, G. Change in View. The M.I.T. Press, 1986.

Harris, Z. S. "Morpheme Alternants in Linguistic Analysis", Language, 1942, 18, 169-180. Reprinted in Joos, 1957.

Harris, Z. S. "From Morpheme to Utterance." Language, 1946, 22, 169-180. Reprinted in Joos, 1957.

Harris, Z. S. Structural Linguistics. University of Chicago Press, 1951.

Harris, Z. S. "Co-occurrence and Transformation in Linguistic Structure." Language, 1957, 33, 283-340. Excerpts reprinted in Syntactic Theory I: Structuralist, edited by F. W. Householder. Penguin, 1972.

Heider, E. Rosch. "'Focal' Color Areas and the Development of Color Names." Developmental Psychology, 1971, 4, 474-455.

Heider, E. Rosch. "Universals in Color Naming and Memory." Journal of Experimental Psychology, 1972, 93.

Heider, E. Rosch, and Olivier, D. C. "The Structure of the Color Space in Naming and Memory for Two Languages." Cognitive Psychology, 1972, 3, 337-354.

Hempel, C. G. Aspects of Scientific Explanation. The Free Press, 1965.

Hirst, R. J. "Phenomenalism." In The Encyclopedia of Philosophy, edited by P. Edwards. Macmillan Publishing, 1967.

Hockett, C. "A System of Descriptive Phonology." Language, 1942, 18. Reprinted in Joos, 1957.

Holbern, H. "Wilhelm Dilthey and the Critique of Historical Reason." In European Intellectual History Since Darwin and Marx, edited by W. W. Wagar. Harper and Row, 1966.

Holdcroft, D. Words and Deeds. Oxford University Press, 1978.

Hollis, M., and Lukes, S., eds. Rationality and Relativism. The M.I.T. Press, 1982.

Hornstein, N. Logic as Grammar. The M.I.T. Press, 1984.

Horowitz, E. and S. Sahni. Fundamentals of Data Structures in Pascal. Computer Science Press, 1984.

Jackendoff, R. Semantic Interpretation in Generative Grammar. The M.I.T. Press, 1972.

Jackendoff, R. X-Bar Syntax: A Study of Phrase Structure. The M.I.T. Press, 1977.

Jackendoff, R. Semantics and Cognition. The M.I.T. Press, 1985.

Jackendoff, R. "Conceptual Semantics." In Meaning and Mental Representations, edited by U. Eco, M. Santambrigio, and P. Violi. Indiana University Press, 1988.

Jackson, P. C. Introduction to Artificial Intelligence. Petrocelli/Charter, 1974.

Johnson, M. The Body in the Mind: The Bodily Basis of Reason and Imagination. University of Chicago Press, 1987.

Joos, M. ed. Readings in Linguistics I. University of Chicago Press, 1957.

Kalish, D. "Semantics." In The Encyclopedia of Philosophy. Volume 7. Edited by P. Edwards. Macmillan Publishing, 1967.

Katz, J. J. "An Outline of Platonist Grammar." In Talking Minds, edited by T. G. Bever, J. M. Carroll, and L. A. Miller. The M.I.T. Press, 1984.

Katz, J. J., and Fodor, J. A. "The Structure of a Semantic Theory." Language, 1963, 39, 170-210. Reprinted in Readings in the Psychology of Language, edited by L. A. Jakobovits and M. S. Miron. Prentice-Hall, 1967.

Kim, J. "Supervenience and Nomological Incommensurables." American Philosophical Quarterly, 1978, 15, 149-156.

Kripke, S. "Semantical Considerations on Modal Logic." Acta Philosophica Fennica, 1963, 16, 83-94. Reprinted in Reference and Modality, edited by L. Linsky. Oxford University Press, 1971.

Kuhn, T. S. The Structure of Scientific Revolutions. 2nd Edition. The University of Chicago Press, 1970.

Lacan, J. The Language of the Self. Tr. by A. Wilden. 1968, reprint Dell Publishing, 1975.

Lakoff, G. "Cognitive Semantics." In Meaning and Mental Representations, edited by U. Eco, M. Santambrogio, and P. Violi. Indiana University Press, 1988.

Lantz, D. and Stefflre, V. "Language and Cognition Revisited." Journal of Abnormal and Social Psychology, 1964, 69, 473;

Leech, G. Semantics. Penguin, 1974.

Lenneberg, E. H. "Cognition in Ethnolinguistics." Language, 1953, 29, 463-471. Reprinted in Language In Thinking, edited by P. Adams. Penguin Books, 1972.

Lenneberg, E. H. "Color Naming, Color Recognition, Color Discrimination: A Reappraisal." Perceptual and Motor Skills, 1961, 12, 375-382.

Lenneberg, E. H. Biological Foundations of Language. John Wiley and Sons, 1967.

Lenneberg, E. H., and Roberts, J. M. "The Language of Experience: A Study in Methodology." International Journal of American Linguistics, 1956, 22, Supplement.

LePore, E., ed. Truth and Interpretation. Basil Blackwell, 1986.

Levelt, W. J. M. Formal Grammars in Linguistics and Psycholinguistics, 3 Volumes. Mouton, 1974.

Lewis, D. Convention. Harvard University Press, 1969.

Lewis, D. "General Semantics." In Semantics of Natural Language, 2nd edition, edited by D. Davidson and G. Harman. D. Reidel, 1972.

Lewis, D. Counterfactuals. Harvard University Press, 1973.

Lewis, D. "Radical Interpretation." Synthése, 1974, 23, 331-344.

Lewis, D. "Languages and Language." In Language, Mind, and Knowledge: Volume VII, Minnesota Studies in the Philosophy of Science, edited by K. Gunderson. University of Minnesota Press, 1975.

Linsky, L. ed. Semantics and the Philosophy of Language. University of Illinois Press, 1952.

Lloyd, B. B. Perception and Cognition: A Cross-Cultural Perspective. Penguin Books, 1972.

Loar, B. "Two Theories of Meaning." In Truth and Meaning, edited by G. Evans and J McDowell. Oxford University Press, 1976.

Loar, B. Mind and Meaning. Cambridge University Press, 1981.

Locke, J. An Essay Concerning Human Understanding. Collated and annotated by A. C. Fraser. In two volumes by Dover Publishing, 1957.

Loux, M. J. ed. The Actual and the Possible. Cornell University Press, 1979.

Lycan, W. G. Logical Form in Natural Language. The M.I.T. Press, 1986.

Lyons, J. Introduction to Theoretical Linguistics. Cambridge University Press, 1968.

Lyons, J. Chomsky. Fontana/Collins, 1970.

Lyons, J. Semantics: Volume I. Cambridge University Press, 1977.

MacIntyre, A. "Relativism, Power, and Philosophy." Proceedings and Addresses of the American Philosophical Association, 1985, 5-22. Reprinted in After Philosophy, edited by K. Baynes, J. Bohman, and T. McCarthy. The M.I.T. Press, 1987.

Maclay, H. "An Experimental Study of Language and Non-linguistic Behavior." Southwestern Journal of Anthropology, 1958, 14, 220-229

Maclay, H. "Overview - Linguistics." In Semantics, edited by D. D. Steinberg and L. A. Jakobovits. Cambridge University Press, 1971.

McDowell, J. "Bivalence and Verificationism." In Truth and Meaning, edited by G. Evans and J. McDowell. Oxford University Press, 1976.

Melden, A. I. Free Action. Routledge and Kegan Paul, 1961.

Mill, J. S. A System of Logic. 1843, reprint Hafner Publishing, 1950.

Miller, G. A. and McNeill, D. "Psycholinguistics." In The Handbook of Social Psychology, Vol. III: The Individual in a Social Context, 2nd edition, edited by G. Lindzey and E. Aronson. Addison Wesley, 1969.

Minsky, M. Computation: Finite and Infinite Machines. Prentice-Hall, 1967.

Minsky, M. "Frame-system Theory." In Thinking: Readings in Cognitive Science, edited by P. N. Johnson-Laird, and P. C. Wason. Cambridge University Press, 1977.

Minsky, M. "A Framework for Representing Knowledge." In Mind Design, edited by J. Haugeland. The M.I.T. Press, 1981.

Minsky, M. The Society of Mind. Simon and Schuster, 1985.

Montague, R. Formal Philosophy. Yale University Press, 1974.

Morris, C. "Foundations of the Theory of Signs." In International Encyclopedia of Unified Science, edited by O. Neurath, R. Carnap and C. Morris. University of Chicago Press, 1938.

Mowrer, O. H. "The Psychologist Looks at Language." American Psychologist, 1954, 9, 660-694.

Nagel, E. The Structure of Science. Harcourt, Brace and World, 1961.

Nagel, E. and J. Newman. Gödel's Proof. New York University Press, 1958.

Nagel, T. Mortal Questions. Cambridge University Press, 1979.

Nagel, T. The View From Nowhere. Oxford University Press, 1986.

O'Hear, A. What Philosophy Is. Penguin Books, 1985.

Osgood, C. E. Method and Theory in Experimental Psychology. Oxford University Press, 1953.

Parsons, T. The Structure of Social Action. 2 Volumes. 1937, reprint The Free Press, 1968.

Parsons T. and E. A. Shils. "Values, Motives, and Systems of Action." In Toward a General Theory of Action, edited by T. Parsons and E. A. Shils, eds. Harvard University Press, 1951.

Partee, B. H. "Some Transformational Extensions of Montague Grammar." Journal of Philosophical Logic, 1973, 2, 509-534.

Peterson, R. L. Concepts and Language. Mouton, 1973.

Platts, M. Ways of Meaning. Routledge & Kegan Paul, 1979.

Popper, K. R. The Poverty of Historicism. Harper Torchbooks, 1957.

Postal, P. "Limitations of Phrase Structure Grammars." In The Structure of Language, edited by J. A. Fodor and J. H. Katz. Prentice-Hall, 1964.

Postal, P. Constituent Structure: A Study of Contemporary Models of Syntactic Description. Indiana University Press, 1964.

Putnam, H. "Minds and Machines." In Dimensions of Mind, edited by S. Hook. New York University Press, 1960. Reprinted in Putnam, 1975.

Putnam H. Contribution to the "Symposium on Innate Ideas." In Boston Studies in the Philosophy of Science, Vol. III. The Humanities Press, 1968.

Putnam, H. "Is Semantics Possible?" In Languages, Belief and Metaphysics, edited by H. Kiefer and M. Munitz. The State University of New York Press, 1970. Reprinted in Putnam, 1975.

Putnam, H. "The Meaning of 'Meaning'." In Language, Mind and Knowledge: Minnesota Studies in the Philosophy of Science, Volume VII, ed. by K. Gunderson. University of Minneapolis Press, 1975. Reprinted in Putnam, 1975.

Putnam, H. Mind, Language and Reality: Philosophical Papers, Volume 2. Cambridge University Press, 1975.

Putnam H. "Why Reason Can't Be Naturalized." Synthése, 1982, 52, 1-23. Reprinted in After Philosophy, edited by K. Baynes, J. Bohman, and T. McCarthy. The M.I.T. Press, 1987.

Quine, W. V. O. Word and Object. The M.I.T. Press, 1960.

Quine W. V. O. "The Problem of Meaning in Linguistics." In Quine, 1961.

Quine, W. V. O. "Two Dogmas of Empiricism." In Quine, 1961.

Quine, W. V. O. From A Logical Point of View. 2nd Edition. Harper and Row, 1961.

Quine, W. V. O. "Quantifiers and Propositional Attitudes." The Journal of Philosophy, 1963, 53, 177-187.

Quine, W. V. O. "Epistemology Naturalized." In Ontological Relativity and Other Essays. Columbia University Press, 1969.

Quine, W. V. O. "Five Milestones of Empiricism." In Theories and Things. Harvard University Press, 1981.

Rickman, H. P. "Geisteswissenschaften." In The Encyclopedia of Philosophy, edited by P. Edwards. Macmillan Publishing, 1968.

Rogers, R. "A Survey of Formal Semantics." Synthése, 1963, 15, 17-56.

Rorty, R. Philosophy and the Mirror of Nature. Princeton University Press, 1979.

Rosch, E. "On the Internal Structure of Perceptual and Semantic Categories." In Cognitive Development and the Acquisition of Language, edited by T. Moore. Academic Press, 1973.

Rosch, E. "Linguistic Relativity." In Human Communication: Theoretical Explorations, edited by A. Silverstein. Lawrence Erlbaum Associates, 1974.

Rosenberg J. F. and C. Travis, eds. Readings in the Philosophy of Language. Prentice-Hall, 1971.

Russell, B. An Inquiry into Meaning and Truth. 1940, reprint Penguin University Books, 1973.

Ryle, G. The Concept of Mind. Penguin University Books, 1949.

Sapir, E. Selected Writings of Edward Sapir in Language, Personality and Culture. Edited by D. G. Mandelbaum. University of California Press, 1949.

Saussure, F. de Course in General Linguistics. 1916, reprint Fontana, 1974.

Scheffler, I. The Anatomy of Inquiry. Bobbs-Merrill, 1963.

Schiffer, S. Meaning. Oxford University Press, 1972.

Schiffer, S. Remnants of Meaning. The M.I.T. Press, 1987.

Schwartz, S. "Introduction." In Naming, Necessity and Natural Kinds, edited by S. Schwartz. Cornell University Press, 1977.

Schwartz, S. ed., Naming, Necessity and Natural Kinds, (Cornell University Press, 1977).

Searle, J. R. Speech Acts. Cambridge University Press, 1969.

Searle J. R., ed. The Philosophy of Language. Oxford University Press, 1971.

Searle, J. "Chomsky's Revolution in Linguistics." The New York Review of Books, 1972. Reprinted in On Noam Chomsky: Critical Essays, edited by G. Harman. Doubleday Anchor, 1974.

Searle, J. R. Intentionality. Cambridge University Press, 1983.

Sells, P. Lectures on Contemporary Syntactic Theories. Center for the Study of Language and Information, 1985.

Shwayder, D. The Stratification of Behavior. Routledge and Kegan Paul, 1965.

Skinner, B. F. Verbal Behavior. Appleton-Century Crofts, 1957.

Slagle, J. R. Artificial Intelligence: The Heuristic Programming Approach. McGraw-Hill, 1972.

Stageberg, N. C. An Introductory English Grammar. Holt, Rinehart and Winston, 1965.

Stalnaker, R. Inquiry. The M.I.T. Press, 1987.

Stefflre, V. C., Vales V. and Morley, L. "Language and Cognition in Yucatán: A Cross-Cultural Replication." Journal of Personality and Social Psychology, 1966, 4, 112-115.

Stillings, N. A., et. al. Cognitive Science. The M.I.T. Press, 1987.

Stitch, S. "Autonomous Psychology and the Belief-Desire Thesis." The Monist, 1978, 61, 573-591.

Stitch, S. From Folk Psychology to Cognitive Science. The M.I.T. Press, 1984.

Strawson, P. F. "Meaning and Truth." In Logico-Linguistic Papers, Methuen, 1971.

Swadesh, M. "The Phonemic Principle." Language, 1934, 10, 117-129. Reprinted in Joos, 1957.

Tarksi, A. "The Semantic Conception of Truth." Philosophy and Phenomenological Research, 1944, 4, 341-376. Reprinted in Semantics and the Philosophy of Language, edited by L. Linsky. University of Illinois Press, 1952.

Tarski, A. "The Concept of Truth in Formalized Languages." In Logic, Semantics, Metamathematics, by A. Tarski. Oxford University Press, 1956.

Vickers, B. ed. Occult and Scientific Mentalities in the Renaissance. Cambridge University Press, 1984.

Wagner, R. The Ring of the Nibelung. Translated by A. Porter.. Faber and Faber, 1976.

Wallace, J. "Only in the Context of a Sentence do Words have any Meaning." In Contemporary Perspectives in the Philosophy of Language, edited by P. A. French, T. E. Uehling, Jr., and H. K. Wettstein. University of Minnesota Press, 1979.

Weber, M. Basic Concepts in Sociology. Translated by H. P. Secher. The Citadel Press, 1962.

Weinstein, S. "Truth and Demonstratives." Nous, 1974, 8, 179-184.

Wells, R. S. "Immediate Constituents." Language, 1947, 23, 81-117. Reprinted in Joos, 1957.

Whorf, B. L. Language, Thought and Reality. Edited by J. B. Carroll. The M.I.T. Press, 1956.

Wierzbicka, A. Semantic Primitives. Athenaum Verlag, 1972.

Wilkes, K. V. Physicalism. Humanities Press, 1978.

Winch, P. The Idea of a Social Science. Routledge and Kegan Paul, 1958.

Winch, P. "Understanding a Primitive Society." American Philosophical Quarterly, 1964, 1, 307-324.

Wittgenstein, L. Philosophical Investigations. 1953, reprint Basil Blackwell, 1976.

Wittgenstein, L. Remarks on the Foundations of Mathematics. 1956, reprint The M.I.T. Press, 1967.

Ziff, P. "On H. P. Grice's Account of Meaning." Analysis, 1967, 28, 1-8. Reprinted in J. F. Rosenberg and C. Travis, eds., 1971.

APPENDIX 1

THE RESEARCH ON THE COLOR DOMAIN

Section 1.5 of Chapter One consists of a brief argument rejecting the claim made by Eleanor Rosch and others that the Linguistic Relativity Hypothesis has been empirically falsified. This appendix presents a review of the research upon which Rosch and others have made attempted to draw these conclusions, and also presents my arguments against the relevance of this research in a little more detailed. This appendix, then, treats the issues of section 1.5 in greater detail.

The Rejection of Whorf's Methodology

Whorf's views became quite influential in the forties and fifties, but a common complaint was that his tantalizing suggestions were extremely difficult to verify empirically. J. A. Fishman published a paper in 1960 in which he presented a scheme designed to distinguish various aspects of the hypothesis, and thereby clarify the problem of empirical verification.[1]

---

[1] J. A. Fishman, "A Systemization of the Whorfian Hypothesis", reprinted in Culture and Cognition, edited by J. W. Berry and P. R. Dasen, (Metheun, 1974), 61-86.

Fishman suggested that the "independent variable" of the Linguistic Relativity Hypothesis, that is, linguistic properties, could be dichotomies into lexical structure and syntactic structure. The first category refers to the referential structure of the language, the second to the organization of linguistic elements. This later category would include Whorf's notion of the underlying semantic structure of the grammatical categories of a language.

Similarly, Fishman has dichotomies the "dependent variable" of the Linguistic Relativity Hypothesis. The dependent variable, according to Fishman, is "cognitive behavior", which he dichotomizes into behavior which is inferred from linguistic data, and behavior which is inferred from non-linguistic data. The former category refers to those properties of cognition which are adduced from an analysis of the linguistic utterances produced by individuals. Whorf's "cosmic forms" are an example of a cognitive feature that is so adduced. The second pole of the dichotomy refers to those properties and features of cognition which can be adduced from behavior that can be observed independently of linguistic considerations. Given these two distinctions, Fishman has identified four "levels" of the Linguistic Relativity Hypothesis:

Figure A.1.1
Fishman's Systemization of the Whorfian Hypothesis[2]

		Data of (cognitive) behavior	
		Language Data (cultural themes)	Non-linguistic data
Data of language characteristics	Lexical characteristics	Level 1	Level 2
	Grammatical characteristics	Level 3	Level 4

In terms of Fishman's systemization, Whorf's primary interest was level 3. Whorf's comments on the effect of language on thought processes and perception (that is, psychological processes which could conceivably be studied independently of linguistically based data) are relatively scant and difficult to interpret. Thus, Whorf does little more than to suggest levels 2 and 4 of the hypothesis. However, it is a historical fact that the majority of social scientists during the forties, fifties, and sixties rejected levels 1 and 3 on methodological grounds, and they insisted that the hypothesis can only be investigated scientifically at levels 2 and 4. One of the most influential advocates of this view was Eric Lenneberg. His 1953 paper

---

[2] Ibid., 83.

"Cognition in Ethnolinguistics"[3] is often referred to as the definitive refutation of levels 1 and 3 of the Linguistic Relativity Hypothesis. It will, of course, be appreciated from the publication dates that Lenneberg did not employ Fishman's terminology in his paper. However, the object of Lenneberg's criticisms is quite clearly the employment of linguistic data as evidence of a linguistic effect on cognition.

Lenneberg characterized Whorf's method of determining cognitive facts as the "translation method in ethnolinguistics."[4] That is, Whorf would often provide a translation of an utterance from an exotic language, such as Hopi, and then proceed to compare the odd sounding English translation with the English sentence that might be uttered in the same situation. Not surprisingly, Whorf often noted significant differences, which he used as evidence for his thesis. Lenneberg found fault with this method on a number of counts. First, Whorf typically would provide translations of the terms of the exotic utterances, and then present the synthetic meaning of the utterance as simply a string of meanings in the same order as the terms of the utterance. Lenneberg showed the inadequacy of this method of translation by demonstrating that English sentences interpreted by this method also come out sounding very different than their normal interpretations. In other words, Lenneberg showed that the method doesn't work for homophonic paraphrase, so there is no reason to believe that it is an adequate method of translation.[5]

---

[3] E. H. Lenneberg, "Cognition in Ethnolinguistics", Language, 1953, 29, 463-471. Reprinted in Language in Thinking, edited by P. Adams (Penguin Books, 1972)

[4] Ibid., 159.

Lenneberg also criticized Whorf's method of translation because it did not appreciate the "metaphorical" element in language. Whorf's approach is based on the assumption that everything said by the speaker of an exotic language must be taken literally. Eleanor Rosch has demonstrated how even a translation of French on this basis can result in the attribution of exotic world views to the speakers of French.[6]

These criticisms of Whorf's method are obviously valid. Lenneberg concludes that the cognitive variation that Whorf claims to have demonstrated is actually just an artifact of his faulty method. Lenneberg then goes on to generalize this conclusion, and claims that the translation between languages of "totally different cultures" is always inexact; always only a "very rough approximation of what has been said and intended originally."[7] Consequently, the cognitive data derived from employing the method will always be artifacts of the method. Lenneberg argues that the translation method must be abandoned in any serious discussion of the Linguistic Relativity Hypothesis, and the cognitive facts must be determined independently of linguistic evidence. The researcher must turn to the non-linguistic behaviors that are "indicative of memory, recognition, learning, problem solving, concept formation and perception." Reiterations of Lenneberg's position were common in the literature at the time.[8] In effect, this amounts to a rejection of levels 1 and 3 as issues that are capable of scientific investigation.

---

[5] Ibid., 160.

[6] E. Rosch, "Linguistic Relativity", Human Communication: Theoretical Explorations, edited by A. Silverstein, (Lawrence Erlbaum, 1974), 97-98.

[7] E. H. Lenneberg, op. cit., 161-162.

The consequence of this widespread rejection of the use of linguistic data in the determination of cognitive facts is that Whorf's original concerns with the total world view of the individual were rejected in favor of the study of the effect of language on specific "cognitive domains." The color domain has been the most has been the most extensively studied of these. In fact, the "domains" that Lenneberg thinks are suitable for investigating the Linguistic Relativity Hypothesis are just those domains that can be operationalized by the methods of psychophysics. However, the concern with specifying cognitive domains that can be operationalized independently of language has had a consequence on the nature of the linguistic variables that were studied. That is, by restricting attention to cognitive domains like color, the syntactic element of language is rendered irrelevant. The restriction to domains that have been operationalized by the methods of psychophysics means that our linguistic concerns are limited to that which Lenneberg calls

---

[8] See, for example, G. A. Miller and D. McNeill, "Psycholinguistics", The Handbook of Social Psychology, Vol. III: The Individual in a Social Context, 2nd edition, edited by G. Lindzey and E. Aronson, (Addison Wesley, 1969), 730-732; and E. Rosch, op. cit.

"the language of experience",[9] that is, language which refers to these cognitive domains. Lenneberg writes:

> The language of experience is particularly well-suited for research because its referents have four advantages over the referents of most other types of words, first, they may be ordered by objective logical criteria (for example, the centigrade scale for temperatures, the frequency scale for pure tones of equal intensity, etc.), whereas furniture, relatives, or most other concrete referents have not such logical and unique orders. Second, the referents have continuity in nature (for example, within certain limits any degree of temperature may be encountered, or any sound frequency heard, whereas the domain of chairs do not grade into the domain of benches, nor do uncles grade into aunts. Third, words in the language of experience refer to closed classes; our sensory thresholds set limits to perception and thus there is a bound to the range of phenomena that may be called hot, loud, green, etc. Fourth, the referents are simple in the sense that each instant may be completely specified by a fixed and very small number of measurements. Temperature is specified by just one measurement; pure tones by two - intensity and frequency; colors by three - for example the Munsell scales of hue, brightness and saturation.[10]

Obviously, the language of experience consists merely of terms and a referential structure. Syntax is irrelevant to the language of experience. Thus, although Lenneberg has started out with the intention of specifying methodological principles that would restrict the social scientist to levels 2 and 4 of the Linguistic Relativity Hypothesis, we see that his argument in effect rules out level 4 as well. Lenneberg's strictures have not been ignored by the researchers; a large majority of the research addressed the Linguistic Relativity Hypothesis has been restricted to level 2, and specifically to the language of experience.

---

[9] E. H. Lenneberg and J. M. Roberts, "The Language of Experience: A Study in Methodology", International Journal of Linguistics, 1956, 22, Supplement.

[10] E. H. Lenneberg, Biological Foundations of Language, (John Wiley and Sons, 1967), 337.

In fact, the very few studies that have been performed on level 4 of the hypothesis have all produced equivocal data.[11] The methodological difficulties and inconclusiveness of this work has been commented on in many places.[12] The level 2 research on the domain of color, on the other hand, has developed over two decades, and according to some of the prominent researchers in the area, some definite conclusions can be drawn on the basis of this research. Consequently, in the next section of this appendix this research will be reviewed and the conclusions based on it will be assessed. Since the few studies that have been performed at level 4 are inconclusive and have not developed into any line of research (and consequently, have not formed the basis of any assessment of the Linguistic Relativity Hypothesis) they will not be discussed.

---

[11] See J. B. Carroll and J. B. Casagrande, "The Function of Language Classification in Behavior", Readings in Social Psychology, 3rd edition, edited by E. E. Maccoby, T. M. Newcomb, and E. L. Hartley, (Holt, Rinehart, and Winston, 1958); H. Maclay, "An Experimental Study of Language and Non-linguistic Behavior", Southwestern Journal of Anthropology, 1958, 14, 220-229; and H. Furth, "The Influence of Language on the Development of Concept Formation in Deaf Children", Journal of Abnormal and Social Psychology, 1961, 63, 386-389.

[12] See D. Lantz and V. Stefflre, "Language and Cognition Revisited", Journal of Abnormal and Social Psychology, 1964, 69, 473; G. A. Miller and D. McNeill, op. cit., 733-735; B. B. Lloyd, Perception and Cognition: A Cross-Cultural Perspective, (Penguin Books, 1972); and E. Rosch, op. cit.

The Color Domain Research

There are a number of reviews of the research inspired by the Linguistic Relativity Hypothesis that are readily available,[13] consequently, the review presented here is not an attempt at complete coverage. The purpose of this section is to present only one line of research, albeit the most extensive line, and to assess the conclusions that some authors have made on the basis of the evidence generated by this research.

The seminal study on the influence of language on cognition was reported by Brown and Lenneberg in 1954.[14] The main independent variable in this study was "codability", a measure of lexical structure. The dependent variable was "recognition", a measure of the accuracy with which a subject could recognize a previously encountered stimulus. The hypothesis was that codability would be positively correlated with recognition. This formal hypothesis was meant to reflect the intuitive idea that a stimulus which is directly associated with a short, common linguistic term would be readily "available" in the cognitive system of the individual. Thus, high codability would mean that any cognitive processing of the stimulus, such as recognition, would be more easily accomplished than would the processing of a stimulus with low availability.

---

[13] See R. W. Brown and E. H. Lenneberg, "Studies on Linguistic Relativity", Readings in Social Psychology, 3rd edition, edited by E. E. Maccoby, T. M. Newcomb, and E. L. Hartley, (Holt, Rinehart, and Winston, 1958); E. H. Lenneberg, Biological Foundations of Language, (John Wiley and Sons, 1967), 337-363; G. A. Miller and D. McNeill, op. cit., 728-750; B. B. Lloyd, op. cit., 36-44; and E. Rosch, op. cit.

[14] R. W. Brown and E. H. Lenneberg, "A Study of Language and Cognition", Journal of Abnormal and Social Psychology, 1954, 49, 454-462.

In order to present Brown and Lenneberg's operational definitions it is first necessary to characterize the "domain" of their experiment and the stimulus materials that were used. The domain is color, and in the field of sensory psychology the domain is viewed as a three dimensional color "solid"; the three dimensions being hue, brightness, and saturation. Any color discrimination made by a human can be represented as a point in this color solid.

A human subject with normal vision can make very fine discriminations between colors; one source suggests that the color solid is differentiatable into 7,500,000 just noticeable differences.[15] It is assumed by most researchers that all normal human beings will make the same set of discriminations, and that this universality is based on the human visual system which is common to people of all cultures and races.

On the other hand, the terms that mean apply to different regions of the color solid vary considerably from language to language. These terms vary not only with respect to their phonetic properties, but also in terms of the regions of the color solid that are named - or not named. (Later we will see that this assumption has been challenged, with drastic consequences for the interpretation of the color domain experiments. But Brown and Lenneberg accepted the assumption, as did most other social scientists at the time. The presentation of Brown and Lenneberg's experiment will be facilitated if we ignore this complication for the time being, and temporarily accept the assumption.) In other words, there is interlinguistic variation in the mapping of terms onto the color solid; that is, there is variation in the lexical structure of the domain. Furthermore, even if we confine ourselves to a single language, it is possible to enquire how "well-coded" the various points of the color solid are. For example, the color of a stop sign is well-coded in English ('red'), whereas the color of oatmeal porridge is less so.

---

[15] Ibid., 457.

In order to operationalize the concept of "well-codedness", Brown and Lenneberg investigated five possible measures of codability: (1) the number of syllables in the naming response to a color, (2) the number of words in the naming response, (3) the reaction time between the presentation of a color stimulus to a subject and his naming response, (4) a measure of intersubjective agreement on naming a color, and (5) a measure of intrasubjective agreement on naming a color. They found a fair degree of correlation between these measures, and finally selected the fourth measure as the operational definition of codability.

Brown and Lenneberg did not, of course, produce codability measures for every just noticeable difference in the color solid. Rather, they produced measures for 24 Munsell color chips,[16] each of which represented a color at maximum saturation. The 24 chips were used in the following experimental procedure. Four of the 24 chips (chosen randomly) were simultaneously presented to a subject and then removed. The subjects were then asked to select the four chips that they had previously observed from an array of 120 chips which included the original 24 chips. It should be noted that all 120 chips were of maximum saturation. Following Lenneberg the 24 chips will be called the 'stimulus colors' and the 120 chip array will be called the 'color context.'[17]

---

[16] Ibid., 459.

The dependent variable of recognition was operationalized by a computation based on the accuracy of response for each of the 24 stimulus colors. This computation was called the recognition score. It was hypothesized that the codability of the stimulus colors would positively correlate with recognition scores. The experimental data supported this hypothesis.

Brown and Lenneberg then investigated the possibility that the recognition scores were due to discriminability rather than codability. This concern was motivated by the knowledge that the 120 Munsell chips of the color context are not perceptually equidistant. The authors suggested that the results might actually be due to the better discrimination conditions of some colors. Additional experimentation ruled out this alternative and supported the original hypothesis (although, again, we will later see that this conclusion was challenged by later researchers). The role of a fourth variable was also investigated. This variable was called the 'storage factor' and it measured the degree to which the cognitive task demands that the stimulus be stored in memory by means of a linguistic term.

---

[17] E. H. Lenneberg, "Color Naming, Color Recognition, Color Discrimination: A Reappraisal", Perceptual and Motor Skills, 1961, 12, 375-382.

In order to determine the role of the storage factor, four experimental groups were subjected to the experimental procedure. The groups varied in the amount of time between the removal of the stimulus colors and the presentation of the color context. The hypothesis was that the correlation between codability and recognition would be strongest for the groups with the highest storage factor, and less strong as the storage factor declined; the reasoning being that the subjects could rely on visual memory for short periods of time. The hypothesis was supported by the data.

Although Brown and Lenneberg's research was intra-cultural, rather than cross-cultural, they suggested that their experiment should be interpreted as support for level 2 of the Linguistic Relativity Hypothesis. They argued that the relation that they found between codability and cognitive processing is an example of a general law which applies to all human beings, now matter what language they speak. But since the lexical structure of the color domain varies greatly from language to language - as the accounts of many anthropologists would have it - the universal law which Brown and Lenneberg discovered would entail quite different consequences for the speakers of different languages.

These cross-cultural assumptions were tested by Lenneberg and Roberts, and the results were published in 1958.[18] The major part of this research involved comparing English to Zuni, the language of an Indian tribe in New Mexico. The specific object of comparison was the referential structure that each language maps onto the color solid. One of the major differences between the two languages is that while English has separate terms for orange and yellow, Zuni has only one term for this entire region of colors. When the Brown and Lenneberg task was administered to Zuni subjects, it was found that compared to English speaking subjects, these subjects were very poor at recognizing the orange and yellow chips.[19] These data support Brown and Lenneberg's claim that the relation between codability and recognition is a universal cognitive law, but because of the interlinguistic variation in the codability of the points of the color solid, different languages facilitate the cognitive processing of particular color stimuli to varying degrees. Again, level 2 of the Linguistic Relativity Hypothesis was supported.

---

[18] E. H. Lenneberg and J. M. Roberts, *op. cit.*

However, in 1961 Lenneberg reported[20] that the results of the original Brown and Lenneberg experiment pertained only to the particular stimulus materials used in that experiment. Lenneberg examined the data of a recognition experiment performed by Burnham and Clark[21] which used a different color context. The array used by Burnham and Clark contained chips that were perceptually equidistant (from their neighbors in the array) but they were not at maximum saturation. By computing codability scores for the colors used in the Burnham and Clark experiment, Lenneberg was able to test the Brown and Lenneberg hypothesis by analyzing the relation between the codability measures and the recognition score data obtained by Burnham and Clark. It was found that with this alternate set of stimulus materials, codability and recognition were negatively correlated. Thus, the Linguistic Relativity Hypothesis was supported only with the maximally saturated color context employed by Brown and Lenneberg, but disconfirmed with the color context employed by Burnham and Clark. The Burnham and Clark color context, incidentally, was another commercially available set of chips known as the Farnsworth-Munsell 100 Hue Test.

---

[19] Ibid., 31.

[20] E. H. Lenneberg, op. cit.

[21] D. Lantz and V. Stefflre, op. cit., 473.

In 1964, Lantz and Stefflre reported and experiment which "attempted to unravel the conflicting relations found between codability and recognition in the studies of Brown and Lenneberg (1954) and Lenneberg (1961) by using a different measure of codability."[22] They suggested that Brown and Lenneberg's measure of lexical structure, i.e., codability, should be replaced with another measure that they called 'communication accuracy.' The rationalization for this new measure was that the relation between language and cognition is that a linguistic term facilitates the cognitive processing of a stimulus because the stimulus can be stored in memory via the linguistic term. Lantz and Stefflre suggested that memory can be metaphorically conceived as a communication process, whereby the individual communicates with himself using the brain as a channel. Moreover, the accuracy with which a stimulus can be communicated intrasubjectively (i.e., the accuracy of memory storage and recall) will be reflected in the accuracy with which the same stimulus can be communicated intersubjectively, using natural language as the medium of communication. Given this argument, the authors concluded that intersubjective "communication accuracy" is a more relevant measure of the "well-codedness" of stimulus than is Brown and Lenneberg's codability measure, which is itself, it will be recalled, operationalized by a computation based on data regarding the intersubjective agreement regarding naming responses to stimuli.

---

[22] Ibid., 473.

In order to operationalize their communication accuracy variable, Lantz and Stefflre asked a group of subjects called 'encoders' to describe a particular color stimuli "in such a way that another person will be able to pick it out."[23] Another group of subjects called the 'decoders' were asked to pick out color chips on the basis of the encoders' descriptions. By comparing the degree of accuracy with which decoders referred back to the stimuli described by the encoders Lantz and Stefflre computed a communication accuracy measure for 20 color chips. These 20 color chips were employed as stimulus colors in a Brown and Lenneberg type experimental procedure. However, two different types of arrays were used as color contexts, one corresponding to Brown and Lenneberg's original experiment (i.e., where the chips were maximally saturated but not perceptually equidistant) and one corresponding to Lenneberg's 1961 data (i.e., where the chips were perceptually equidistant but not maximally saturated). It was hypothesized that there would be a positive correlation between communication accuracy and recognition for both types of array. Furthermore, measures of intersubjective naming agreement and the mean number of words to describe a chip were counted for each chip. It was hypothesized that communication accuracy would be a better predictor of recognition than would either of these two measures. The data supported all of the author's hypotheses, and they concluded that communication accuracy would be positively correlated with recognition no matter what sort of stimulus materials were used.

---

[23] Ibid., 474.

Just as Brown and Lenneberg's intracultural research was replicated cross-culturally, so was Lantz and Stefflre's work replicated cross-culturally by Stefflre, Vales and Morley. In their 1966 report,[24] the authors described a replication of the Lantz and Stefflre experiment, except that only one stimulus array, the Farnsworth-Munsell array, was used. There were two groups of subjects, those who spoke primarily Spanish and those who spoke primarily Yucátec, a Mayan language. Even though particular color chips received different communication accuracy scores in the different languages, a positive correlation between communication accuracy and recognition was found in the data obtained from both groups. The same conclusions can be drawn from this cross-cultural work that was earlier drawn from the work of Lenneberg and Roberts, i.e., the relation between lexical structure and cognitive processes is a universal law that has different effects for different linguistic groups because of the different lexical structures of various languages. Thus level w of the Linguistic Relativity Hypothesis is supported by this research. The only change from the previous conclusion is that communication accuracy has replaced codability as a measure of lexical structure.

---

[24] V. Stefflre, C. V. Vales and L. Morley, "Language and Cognition in Yucatán: A Cross-Cultural Replication", Journal of Personality and Social Psychology, 1966, 4, 112-115.

However, Lenneberg has subsequently argued that the experiments employing communication accuracy as a linguistic variable are irrelevant to an evaluation of level 2 of the Linguistic Relativity Hypothesis.[25] Lenneberg's position is based on an acceptance of the "weakened" principle of linguistic relativity that was expressed in Brown and Lenneberg's 1954 paper. It was argued there that language does not determine what can and cannot be cognitively processed by subjects. Rather, linguistic relativity is taken to mean that lexical structure would determine the relative "availability" of various referents for cognitive processing, and therefore would determine the rate and accuracy with which the referents would be processed. Thus, Brown and Lenneberg accepted the principle that any object, event, etc. that is described in any language is also describable in any other language. Their concept of codability was an attempt to capture the relative "ease" by which referents are coded within a language. Thus, their "weakened" version of the Linguistic Relativity Hypothesis was operationalized as a hypothesis about the relation between codability and cognitive structure.

---

[25] E. H. Lenneberg, Biological Foundations of Language, (John Wiley and Sons, 1967), 337-363.

Lenneberg argues that the concept of communication accuracy, on the other hand, is not a valid measure of lexical structure. Every language can code any particular referent, and furthermore, any referent can be coded by a large number of phrases, for example, 'brown', 'the color of the bricks in Corbett Hall', 'the color of the rust on that car', etc. These descriptive phrases, and the success with which they facilitate communication "will depend frequently on individual ingenuity rather than on the language spoken by the communicator."[26] It is also obvious that the communicative efficiency of these phrases will depend to some degree on the particular experiences of individuals. Consequently, any experiment which uses communicative accuracy as an independent variable confuses the effect of objective lexical structure with idiosyncratic variations in fluency and experience in ways that are impossible to unravel. These latter effects must be excluded in any experiment designed to test the Linguistic Relativity Hypothesis. These problems are serious enough for intracultural research; they are even worse for cross-cultural research, since it may be assumed that the members of different cultures will vary both in the nature of their experiences and in their motivation to provide precise descriptive accounts of particular referential domains. It may be concluded then, that Lenneberg was correct in claiming that communication accuracy is not a measure of the lexical structure of a language, and consequently, it has no relevance to level 2 of the Linguistic Relativity Hypothesis. This means, of course, that the experiments of Lantz and Stefflre and of Stefflre, Vales and Morley are not relevant to the Linguistic Relativity Hypothesis. It also means that the contradictory results in the color task that Lenneberg described in 1961 is still unexplained.

---

[26] Ibid., 355.

The next phase of the research on the color domain has been conducted almost entirely by one person, Eleanor Rosch (formerly Eleanor Rosch Heider; her earlier experiments are published under her former name). Rosch's research suggests a way of reconciling the contradiction discussed above, but it also suggest that the lexical coding of the domain of color has no relativistic effect on the cognitive processing of color stimuli. That is, Rosch's work appears to be a disconfirmation of level 2 of the Linguistic Relativity Hypothesis.

In order to appreciate the direction of Rosch's research it is first necessary to consider a book by Berlin and Kay entitled Basic Color Terms,[27] since Rosch's work incorporates some revolutionary ideas that were advanced by Berlin and Kay. Up to the time of Berlin and Kay's work, anthropologists had investigated the color terminologies of a large number of languages. On the basis of these investigations it was commonly concluded that the lexical coding of the color domain in each language is entirely arbitrary; that is, the regions of the color solid that are named or not named in a particular language has no essential relation to the way that the color solid is coded in any other language. This assumption underlies the cross-cultural studies of Lenneberg and Roberts, and of Stefflre, Vales and Morley. In fact, the universal law relating lexical structure to cognitive processing can only be interpreted as evidence of level 2 of the Linguistic Relativity Hypothesis if, indeed, lexical structures vary arbitrarily.

---

[27] B. Berlin and P. Kay, Basic Color Terms: Their Universality and Evolution, (University of California Press, 1969).

Berlin and Kay have challenged this underlying assumption, and argued that color categorization is not arbitrary, and that the referents of "basic" color terms are similar for all languages. In order to distinguish the "basic" color terms of a language from other utterances whose referents are also colors, Berlin and Kay have described four criteria which a color term must meet if it is to qualify as "basic": (1) It must be monolexemic, i.e., its meaning must not be predictable from the meaning of its parts. (2) Its referent should not be included in any other color term. (3) Its application must not be limited to a narrow class of objects. And finally, (4) it must be psychologically salient for informants.[28] On the basis of these criteria, and a few other criteria designed to handle the odd doubtful case, Berlin and Kay have argued that all 98 languages that they investigated have eleven or fewer basic color terms. They then went on to map the basic color terms of 20 languages[29] of considerable genetic diversity. They used an adaptation of the comparative mapping methodology of Lenneberg and Roberts,[30] which used informant judgements about a display of 320 maximally saturated Munsell chips. The referent of each color term was mapped in two ways by Berlin and Kay; they recorded both the boundary of the color term and the chip that represented the "best example" of the color term. This latter mapping was called the 'category foci placement.' The authors found that the boundary method of mapping determined very different mappings for speakers of different languages, but it also found that there was as much variance between speakers of the same language, and even between the same subject on different trials. It was concluded that the boundary method must be rejected as a mapping technique on the grounds of unreliability. Category foci placements, on the other hand, proved to be a highly reliable intralinguistic mapping technique; category foci were rarely displaced more than two adjacent chips.[31] Furthermore, it was found that the "location of color foci varies no more between speakers of different languages than between speakers of the same language."[32] In other words, when foci placement replaces boundary determination as a method of mapping color terms to the color solid, the mapping is universal to all languages containing a full eleven basic color terms, i.e., for all languages with eleven terms, the eleven terms refer to the same eleven colors. The previous anthropological conclusion on the relativity of color terminologies had been based on an incorrect emphasis on the unreliable technique of mapping color boundaries.

---

[28] *Ibid.*, 6.

[29] *Ibid.*, 6.

[30] Lenneberg and Roberts, *op. cit.*

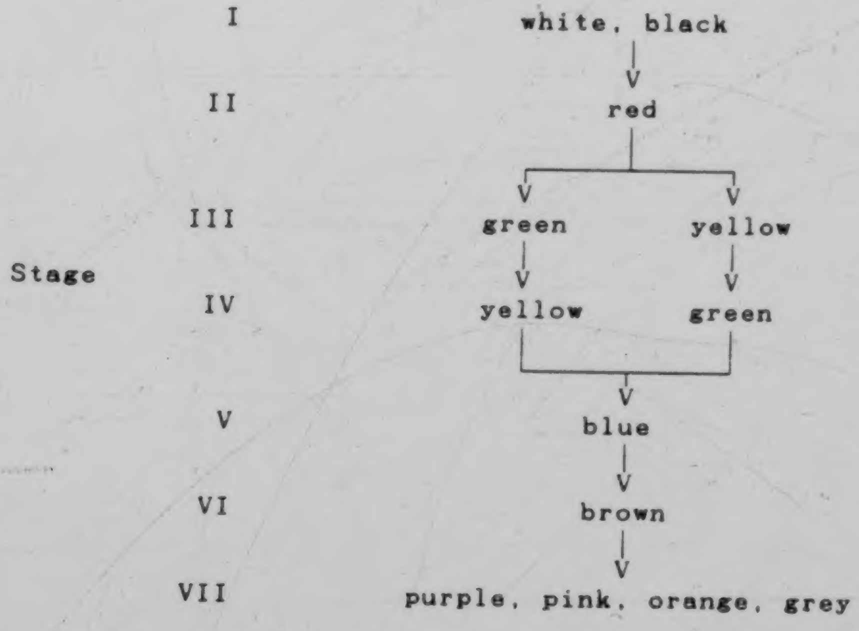
---

[31] Ibid., 13.

[32] Ibid., 10.

What about languages with less than eleven basic color terms? Is the lexical structure of these languages arbitrary with respect to the color domain? It is not, according to Berlin and Kay. They hypothesize that the historical evolution of basic color terms follows a fixed course in every language. All languages contain a term for black and white, and from this base, only a fixed number of evolutionary paths are possible. These paths, and the seven stages in a culture's acquisition of the eleven basic color terms are represented in the figure below.

Figure A.1.2
Berlin and Kay's Model of the Evolution of Color Terms



I
white, black
II
red
III
green
yellow
Stage
IV
yellow
green
V
blue
VI
brown
VII
purple, pink, orange, grey

Consequently, according to Berlin and Kay, if the lexical structure of the color terms of two languages differ it is not because of an arbitrary relativism, rather, it is because4 the two languages are a different stages of a universal evolutionary sequence.

Influenced by these arguments, Rosch presented the following hypothesis:

> ...there are specific areas of the color space - defined as particular intersections of the three dimensions hue, value, and saturation - which are universally most codable and most accurately remembered (both in long term and short term memory) and it is these areas that form the focal points of the basic color names across languages.[33]

In other words, Rosch was suggesting that the results of the Brown and Lenneberg experiment are due to a "natural" system of differential perception and memory, rather than a supposed arbitrary difference in codability imposed by different languages. To put the same point another way, Brown and Lenneberg had assumed that all the points in the color solid have an equal perceptual significance, whereas Rosch is claiming that some colors, the colors corresponding to Berlin and Kay's eleven basic color terms, have a greater perceptual saliency than other regions of the color solid.

The data collected by Berlin and Kay did not constitute proof of the proposition that focal colors are more salient, and therefore more codable, so Rosch had to design her own experiments. She first extended Berlin and Kay's determination of the focal points of the basic color terms to include the saturation dimension of the color solid. Berlin and Kay's stimulus materials, it will be recalled, consisted of 320 chips, all of which were at maximum saturation. Rosch investigated the possibility that some non-saturated focal points may exist. However, she found that the original assumption of Berlin and Kay was correct; all the focal points of the basic color terms of the eleven different languages were at maximum saturation.[34]

---

[33] E. Rosch Heider, "Universals in Color Naming and Memory", Journal of Experimental Psychology, 1972, 93, 11.

After determining the focal points of the basic color terms of several languages, and confirming that these points were the same for all of the languages tested, Rosch compared the codability of focal and non-focal colors in several languages. (Henceforth "focal colors" will refer to colors which are the focal points of the eleven basic color terms. "Non-focal colors are all other points in the color solid.) The results were that codability, measured by Brown and Lenneberg's first three measures (cf. above) were significantly higher for focal colors than for non-focal colors.[35] On the basis of these experimental results, Rosch has provided an interpretation of the contradiction between Brown and Lenneberg 1954 data and the Burnham and Clark data that Lenneberg reported in 1961. It will be recalled that the Burnham and Clark experiment used the Farnsworth-Munsell 100 Hue Test as stimulus materials, and this array does not include any chips at maximum saturation. In other words, there are no focal points of basic color terms in the farnsworth-Munsell stimulus materials. Rosch claims that it is this fact which accounts for the lack of a positive correlation between codability and recognition in the Burnham and Clark experiment. The Brown and Lenneberg experiments, on the other hand, employ stimulus materials which contain the focal colors, and it is the salience of these focal colors, rather than codability, which is responsible for enhanced recognition memory.

---

[34] Ibid., 13.

[35] Ibid., 14.

Rosch's interpretation raises a question however. Why is it that the previous cross-cultural studies as level 2, i.e., the 1956 studies of Lenneberg and Roberts and the 1966 studies of Stefflre, Vales and Morley, did not reveal the universality of color naming and memory? The authors of these studies concluded just the opposite, that there is a difference between the lexical structure of the languages studies, and that this difference in lexical structure is reflected in the memory performance of the subjects. Rosch has discussed the Stefflre, Vales and Morley study from this point of view. She argued that since this study employed the Farnsworth-Munsell array, there were no focal colors in the stimulus materials. Since the basis of the universality claim is that the focal colors are perceived and cognitively processed in a universal fashion, it is not to be expected that Stefflre, Vales and Morley should have found the same memory performance in the different cultural groups.[36]

Rosch does not, however, provide any explanation of why the Lenneberg and Roberts data turned out as it did. Lenneberg and Roberts, it will be recalled, employed and array which included focal colors, so according to Rosch's universality hypothesis, the lexical structure of Zuni should be the same as English.

---

[36] *Ibid.*, 19.

Brown and Lenneberg[37] insisted that the yellow-orange region of the color solid is coded with a single term in Zuni. However, the Lenneberg and Roberts report indicates that there are separate Zuni terms for orange and yellow. This contradiction is not resolved in the papers cited. There is also a contradiction between Lenneberg and Roberts and Berlin and Kay on the placement of the basic color terms of Zuni. About the only conclusion one can draw from this confusion is that Rosch has taken Berlin and Kay's position on this issue, and has argued that the focal points of the basic color terms of English and Zuni are identical.

In any case, Rosch has produced a number of other experiments which tend to support her claim that the perceptual saliency of the focal colors underlies both codability and cognitive processing. One set of three experiments was performed on 3 and 4 year-old American children.[38] In experiment I, 3 year-olds were asked to pick any color they chose from an array of chips which included both focal and non-focal colors. No color terminology was employed by the experimenters in this task. It was found that children chose focal colors considerably more often than non-focal colors, and this was interpreted as indicating that the focal colors, because of their perceptual salience, attracted the attention of the children. Experiment II was a matching task. Subjects were shown a chip and asked to point to the matching chip in a simple one-dimensional array. Again, no color terminology was employed by the experimenters. Accuracy scores for focal colors were found to be significantly higher than for those of non-focal colors. This was again interpreted as evidence that the focal colors were more perceptually salient for the children. Finally, experiment III was designed to indirectly test the hypothesis that children learn color names by first attaching names to the perceptually salient focal colors, and only later do they expand their terminology from this base. Since the 3 and 4 year-old children already possessed color vocabularies, this hypothesis had to be replaced with the related hypothesis that the young subjects will identify focal color chips as the best examples of basic color terms. The eight chromatic basic color terms (i.e., excluding 'black', 'white', and 'grey') of English were employed in the experiment. It was found that if obviously incorrect choices were excluded the data showed that the children chose focal chips significantly more often than non-focal chips.

---

[37] See R. W. Brown and E. H. Lenneberg, "A Study in Language and Cognition", Journal of Abnormal and Social Psychology, 1954, 49, 454-462; and R. W. Brown and E. H. Lenneberg, "Studies in Linguistic Relativity", Readings in Social Psychology, 3rd edition, edited by E. E. Maccoby, T. M. Newcomb, and E. L. Hartley, (Holt, Rinehart and Winston, 1958).

[38] E. Rosch Heider, "'Focal' Color Areas and the Development of Color Names", Developmental Psychology, 1971, 4, 474-455.

Experiment III does lend some support to Rosch's hypothesis about the ontogenesis of color names, but it is clear that an experiment of this type cannot provide very strong evidence for the hypothesis. What is needed is a group of subjects who do not possess a color vocabulary so that their original acquisition of color terms can be studied. Children within our own culture provide a possible source of such subjects, but a longitudinal study of the natural acquisition of color terms is a difficult undertaking. The artificial laboratory teaching of color terms is also ruled out because of the cognitive limitations and motivational characteristics of young children. However, another potential source of subjects is from the adult population of linguistic communities whose languages have very few basic color terms, i.e., languages which are at the early stages in Berlin and Kay's proposed evolutionary sequence of color terms. One such linguistic community is the Dani of New Guinea. These people have a Stage I color terminology, i.e., they have only two color terms. These terms correspond roughly to 'black' and 'white', but perhaps a better English translation would be 'light' and 'dark'. In any case, Rosch has tested her developmental hypothesis in experiments where Dani were taught the color terms under controlled conditions.[39]

The first experiment with Dani subjects described by Rosch was a replication of the Brown and Lenneberg experiment with an English speaking group and a Dani group (as yet untrained in names for chromatic colors). Rosch hypothesized that the Dani group would exhibit memory differences between the focal and the non-focal colors, even though the codability differences do not exist in the Dani language. The results showed that the American subjects were more accurate than the Dani in the recognition of both focal and non-focal colors. But within each linguistic group the differences between the recognition of focal and non-focal colors were identical. Both American and Dani subjects had higher recognition accuracy scores with focal stimulus colors than with non-focal stimulus colors.[40]

---

[39] See experiments III and IV from E. Rosch Heider, "Universals in Color Naming and Memory", Journal of Experimental Psychology, 1972, 93; and experiment I from E. Rosch, "On the Internal Structure of Perceptual and Semantic Categories", Cognitive Development and the Acquisition of Language, edited by T. E. Moore, (Academic Press, 1973).

Rosch then tested the developmental hypothesis directly with Dani subjects. An experiment was designed which employed color chips for the eight chromatic focal colors, and another eight chips which represented the non-focal colors. The Dani subjects were paid to learn 16 stimulus-response pairs. A color chip was the stimulus and a neutral Dani word (the words employed were the names of Dani descent groups) was the response. Each subject had to learn a different set of color-term associations in order to minimize any uncontrolled effects in the memorability of the terms. The procedure employed in the learning situation was that the subject was first told the names of the 16 chips, and were then asked to repeat the names when the chips were shown in random order. Each subject received five runs a day, every second day, until a perfect run was obtained. A average of 3.5 days was required to obtain a perfect run.

The hypothesis was that the focal colors, because of their saliency, are more easily retained in long term memory and consequently can more easily become associated with names. The results showed that the mean number of error for the eight non-focal colors was significantly higher than for the focal colors, thus supporting the hypothesis. A secondary hypothesis was also tested. It was suggested that the order of difficulty of learning the focal colors would correspond to the evolutionary order of color terms proposed by Berlin and Kay, with the easiest-to-learn colors being those that come earliest in Berlin and Kay's sequence. The actually obtained rank order of difficulty did not significantly correspond to Berlin and Kay's sequence.[41]

---

[40] E. Rosch Heider, "Universals in Color Naming and Memory", Journal of Experimental Psychology, 1972, 93.

The results of the previous experiments all supported Rosch's developmental hypothesis. The developmental hypothesis suggests, at first glance, that the causal direction specified by level 2 of the Linguistic Relativity Hypothesis is actually just the reverse. I.e, it is cognitive-perceptual mechanisms which determine lexical structures rather than the other way around. However, it must be recalled that Rosch's developmental hypothesis applies only to the focal colors. What if we consider only colors of less than maximal saturation; does the arbitrary lexical structure mapped onto these colors affect the cognitive processing of these color stimuli? It may be the case that Rosch's developmental hypothesis correctly characterizes the relation between lexical structure and cognition for the focal colors, but the Linguistic Relativity Hypothesis correctly characterizes the lexical-cognitive relations for the non-focal colors. It will be recalled that some of the previous research has used the Farnsworth-Munsell array which does not contain focal colors, and consequently, this research might be relevant to this "restricted" Linguistic Relativity Hypothesis. Lenneberg's 1961 analysis of Burnham and Clark's data showed that codability was negatively correlated with recognition memory, thus the restricted Linguistic Relativity Hypothesis was not supported. The research of Lantz and Stefflre also used the Farnsworth-Munsell array, but the independent variable was communication accuracy. As Lenneberg has argued,[42] this variable is not a valid measure of lexical structure, and therefore this experiment is no relevant to the Linguistic Relativity Hypothesis, even in the "restricted" form being discussed here. Consequently, what evidence existed tended not to support the restricted Linguistic Relativity Hypothesis.

---

[41] Ibid., 18, 19-20.

Rosch designed an experiment[43] which was intended to bring further evidence to bear on this restricted form of the Linguistic Relativity Hypothesis. The measures that were employed in this experiment were rather elaborate; measures where constructed for the "structure" of memory and of naming. The domain over which these measures were computed was a set of less than fully saturated color chips; in effect, the focal colors were excluded. Naming and memory structures were computed for two groups of subjects, Americans and Dani who knew only the Dani two color vocabulary. Each group contained about equal numbers of children and adults. The experiment was designed to test the restricted Linguistic Relativity Hypothesis with the following specific hypothesis: within each linguistic group the naming structure and the memory structure will be to some extent isomorphic, but between the linguistic groups the structures will not be similar.

---

[42] E. H. Lenneberg, The Biological Foundations of Language, (John Wiley and Sons, 1967).

[43] E. Rosch Heider and D. C. Olivier, "The Structure of the Color Space in Naming and Memory for Two Languages", Cognitive Psychology, 1972, 3, 337-354.

All the subjects had to perform in a naming task and a memory task. In the naming task subjects were asked to name each separate color of a 40 member stimulus array. In the memory task, subjects were shown a single stimulus for 5 seconds, then after a 30 second unfilled interval they were asked to identify the exposed chip from a 40 chip array, the same one used in the naming task. The data obtained from these tasks were displayed in four matrices, as described in the following passage:

> Separate 40 X 40 similarity matrices were constructed for the memory and naming data for each color. In the memory matrix the (i,j) entry was the number of times chip i elicited chip j as a response. In the naming matrix the (i,j) entry was the number of Ss who gave chip i and chip j the same name; an S was counted whenever his name was identical word for word to his name for chip j.[44]

Multidimensional scaling techniques were applied to these matrices, and as a result, each matrix was transformed into a set of 40 points in three dimensional Euclidean space, each of the points corresponding to a chip in the original stimulus array. The distance between chips can be interpreted in the following way: a short distance between chips means that they are more likely to be confused in memory/naming, while a long distance means that confusion is less likely. Graphs were drawn of four structures: Dani color naming, American color naming, Dani color memory, and American color memory. Visual examination of these graphs shows that while the Dani and American color naming structures are dissimilar, the color memory structures are very similar, which suggests that the restricted Linguistic Relativity Hypothesis was not supported. In order to obtain a less subjective comparison of the four structures, Schonemann and Carroll's measure of departure from good fit was employed. Measures of departure from good fit, i.e., measures of dissimilarity of structure, were as follows:[45]

---

[44] Ibid., 343.

Dani naming versus American naming	.194
Dani memory versus American memory	.161
Dani naming versus Dani memory	.126
American naming versus American memory	.212

These figures support the visual impression that Dani and American memory structures are more alike than their naming structures (.161 compared to .194). However, the figures also indicate that the two most similar structures are Dani naming and Dani memory, rather than the two memory structures. (Incidentally, this conflicts with the visual impression obtained from viewing the graphs, as the reader can confirm by consulting the original paper.) In any case, the large difference between the American naming and memory structures means that the hypothesis of the experiment is not supported. This suggests that even if level 2 of the Linguistic Relativity Hypothesis is restricted to non-focal colors, it must be rejected as an account of the relation between the lexical coding and the cognitive processing of the color domain.

---

[45] Ibid., 348.

Rosch has continued with her studies of the Dani; she has reported an experiment which suggests that color categories are "internally structured" in the sense that they have focal areas and surrounding non-focal areas, and that the key to learning the term referring to a category is the initial attachment of the term to the focal area on the basis of the perceptual saliency of the latter.[46] From the point of view of the Linguistic Relativity Hypothesis, these experiments do not represent any advance over the position stated in the paper on naming and memory structures, but they do provide additional evidence for Rosch's developmental hypothesis. Along with the previous experiments that have been reviewed, there seems to be little experimental support for the Linguistic Relativity Hypothesis when the domain is color. These experiments and their relevance for level 2 of the Linguistic Relativity Hypothesis are summarized in the table below.

Table A.1.1
Summary of the Color Domain Experiments

Brown and Lenneberg (1954)	Interpreted by the authors as intracultural evidence in support of the LRH, but Heider (1972) argued that codability is related to the perceptual saliency of the focal colors, therefore the data are equivocal.

---

[46] E. Rosch, "On the Internal Structure of Perceptual and Semantic Categories", Cognitive Development and the Acquisition of Language, edited by T. E. Moore, (Academic Press, 1973).

Lenneberg and Roberts (1956)	Interpreted by the authors as cross-cultural evidence of the LRH, but Berlin and Kay (1969) and Heider (1972) have questioned the assumptions behind the corss-cultural comparisons presented in this report.
Lenneberg (1961)	Compared the data of Brown and Lenneberg (1954) and Burnham and Clark (1955), and found contradictory relations between the linguistic and cognitive variables. This was attributed to differences in the arrays used: Brown and Lenneberg used only maximally saturated chips, where Burnham and Clark used the Farnsworth-Munsell array. This discrepancy was later discussed by Lantz and Stefflre (1964) and Heider (1972).
Lantz and Stefflre (1964)	Attempted to resolve the discrepancy identified by Lenneberg (1961) by constructing a new linguistic variable, communication accuracy. The authors concluded that with the new variable, their work provided intracultural evidence for the LRH. Lenneberg (1967) argued that this experiment is not relevant to the LRH because communication accuracy is not a valid measure of lexical structure.
Stefflre, Vales and Morley (1966)	Interpreted by the authors as cross-cultural evidence in support of the LRH. Criticized by Lenneberg (1967) for the same reason that he criticized Lantz and Stefflre.
Heider (1972) "Universals" Exp. I and II	Found that focal colors were more codable than non-focal colors. Also found that the focal colors were named with the same lexical structure in a large number of languages. These results were used to reinterpret Brown and Lenneberg's 1954 experiment as not supporting the LRH.
Heider (1972) "Focal Color Areas"	Found that for English speaking children focal colors are more perceptually salient than non-focal colors. From this the author suggested that the learning of basic color terms is mediated by the universal perceptual saliency of focal colors.

Heider (1972) "Universals" Exp. III and IV	This cross-cultural research demonstrated that focal colors are more easily recognized than non-focal colors, and that it is easier to learn the names of focal colors than non-focal colors. Interpreted by the author as evidence against the LRH and in support of her developmental hypothesis.
Heider and Olivier (1972)	Tested a "restricted" version of the LRH, specifically, that lexical structure affect the cognitive processing of only the non-focal colors. This restricted hypothesis was not supported.
Rosch (1973) Exp. I	Further cross-cultural evidence in support of the author's developmental hypothesis.

The history of the research at level 2 of the Linguistic Relativity Hypothesis, along with the failure of any conclusive body of research to materialize at level 4, has lead Rosch to the following conclusion:

> We began with the notion of linguistic relativity defined in terms of insurmountable differences in the world view of cultures brought about by differences in natural languages. Because of the variety of requirements for specificity and cross-cultural controls in testing such assertions, we were reduced to the far less sweeping claim that color names affect some aspects of thought. However, we discovered that color appears to be a domain suited to demonstrate just the opposite of linguistic relativity, namely, the effect of the human perceptual system in determining natural categories... At present, the Whorfian hypothesis not only does not appear to be empirically true in an major respect, but no longer seems profoundly and ineffably true.[47]

In other words, Rosch is arguing that the Linguistic Relativity Hypothesis should be rejected as a plausible account fo the relation between language and cognition. Levels 1 and 3, the latter being the "profound and ineffable" level of the hypothesis, must be rejected as scientific hypotheses because of the impossibility of testing them on an empirical basis. Level 4 must be rejected because it has not been possible to generate any fruitful hypotheses at this level. Finally the research on the color domain conducted at level 2 suggests that the relation between language and cognition might be just the opposite as specified by the Linguistic Relativity Hypothesis.

---

[47] E. Rosch, "Linguistic Relativity", Human Communication: Theoretical Perspectives, edited by A. Silverstein, (Lawrence Erlbaum, 1974), 119.

Is Rosch's conclusion justified? I believe it is not. A minor problem with Rosch's conclusion is that it involves a hasty generalization. That is, she has concluded that the negative verdict from the color research generalizes to all possible research that might be conducted at level 2 or level 4. While this might be true, Rosch provides no reasons that would justify this sweeping generalization.

However, the real problem with Rosch's conclusion is with the initial rejection of level 3 of the hypothesis as "unscientific". It will be recalled that this rejection was first recommended by Eric Lenneberg in 1953 when he argued that Whorf's "translation method" is methodologically flawed, and that to ensure scientific objectivity the researcher must avoid deriving "cognitive" data from "linguistic" evidence. I fully agree that there is something essentially "unscientific" about Whorf's method, and for that matter, any method that attempts to "interpret" the beliefs of a person. However, the shift to level 2 and 4 on the grounds of methodological purity is to throw the baby out with the bathwater. The simple fact of the matter is that Fishman's "systemization" of the Whorfian hypothesis is a misrepresentation. Whorf's hypothesis is nothing more than Fishman's level 3. Levels 1, 2, and 4 may have some independent interest but they are not what Whorf was talking about. Consequently, the research on the color domain does nothing to refute Whorf.

Lenneberg rejected level 3 because it seemed to be based on an "unscientific" method whereby cognitive facts (beliefs) are inferred from data that is in part linguistic. As should be apparent from my chapters three and four, I agree with Lenneberg that such a methodology is, in a sense, unscientific. Where I disagree with Lenneberg is in his optimism that this method can be excised and replaced with the more clinical methods of experimental psychology, while still remaining true to Whorf's intentions. If we what we are interested in is an examination of what Whorf was actually talking about, then we must necessarily investigate the messy, "unscientific" methodology of interpretation, for that method, and not the methods of experimental psychology, is essential to the Linguistic Relativity Hypothesis.

APPENDIX 2

A SUMMARY OF THE ARGUMENT

This appendix provides a summary of the main line of argument used in the body of the thesis.

CHAPTER ONE - WHORF'S HYPOTHESIS

1.1 WHORF'S HYPOTHESIS

Early writers such as Aristotle and Locke held that thought is independent of and prior to the use of language. Whorf argued that this view is incorrect. He argued that our very ability to think is due to our use of language, and moreover, because languages vary, thought will also vary between speech communities. Whorf's view is called the Linguistic Relativity Hypothesis.

1.2 WHORF'S ARGUMENT

Whorf's argument for the Linguistic Relativity Hypothesis consists of the following five theses:

1. To think is to employ concepts.

2. Languages have a conceptual structure.
(That is, a system of concepts is associated with each natural language just as a syntax is associated with each.)

3. Languages differ.
(That is, different natural languages have different conceptual structures.)

4. Language determines thought.
(That is, when we learn a natural language we acquire an unconscious command over its conceptual structure, and this conceptual structure becomes a foundation for all our concepts.)

5. Cross-cultural understanding is difficult/impossible.
(That is, if language A and language B are very different, the concepts of the speakers of language A will be very difficult, if not impossible, to express in language B.)

1.3 THE WESTERN MIND AND THE HOPI MIND

Whorf developed two examples of how language effects thought. One is that the group of languages that he called "Standard Average European" have a tendency toward "objectification", that is, a large number of words behave syntactically and semantically the same way as words referring to physical objects. Thus speakers of these languages tend to view physical objects as fundamental notions, and the notions of space and time in which physical objects exist are likewise fundamental to the speakers' conceptual scheme.

On the other hand, Hopi avoids the linguistic pattern of "objectification" and consequently the fundamental concepts of the Hopi are not space, time and physical objects. Instead, their fundamental concepts are "manifested" and "manifesting", and their experienced world is more a world of process than of things.

1.4 MIND: CONTENT OR CONSCIOUSNESS?

In order to assess Whorf's Linguistic Relativity Hypothesis is necessary to be clear about what we mean by "thought" and "mind". Historically, there have been two distinct approaches to characterizing the nature of mind. One approach, originating largely from Descartes, is that the essence of mind is consciousness (private experience, sentience, etc.). The other approach, often associated with Brentano, is that the essence of mind is its intentionality (that is, mental phenomena is always "about" something, it always has "content").

In this work I will focus solely on mind in Brentano's sense. However, unlike some recent authors, I do not think that Descartes' approach to mind is totally on the wrong track. Thomas Nagel has argued that we require an intellectual breakthrough in order to properly account for mind in Descartes' sense. I think that Nagel may be on the right track, however, in this work I assume that the Linguistic Relativity Hypothesis is claim about how the content of thought varies, so I restrict myself to Brentano's conception of mind.

1.5 EMPIRICAL RESEARCH

Whorf's writings inspired a body of empirical research in the 1950's and 60's. The researchers involved argued that

Whorf's thesis was intolerably vague, and that it had to be reworked into a set of precise empirical hypotheses. The resulting research has been taken, by its originators, as demonstrating the falsity of the Linguistic Relativity Hypothesis. My view is that these empirical hypotheses do not represent a fair translation of Whorf's intentions. Whorf's thesis is an attempt to make a fairly high-level theoretical claim, and it is at that level that I will approach it. The empirical research, while interesting in its own right, is irrelevant to Whorf's thesis.

1.6 PREVIEW

This section is an overview of the following three chapters. In chapter two I will argue that the proper way to study natural languages is to adopt what is called the "truth-conditional program". The basic idea is that we describe a natural language by constructing a Tarskian "truth-definition" for the language (with appropriate modifications). However, this program only makes sense when it is embedded in a larger theory of "action", of the type that Max Weber advocated.

Chapter three argues that "thought" should be understood in terms of propositional attitude psychology, that is, as the dynamics of the beliefs, desires, etc., of the psychological subject. A question is raised: How does propositional attitude psychology "fit" with the natural sciences? I argue (against Jerry Fodor) ~~that~~ there is no fit; that propositional attitude psychology does not conform to the scientific world view. Consequently the Linguistic Relativity Hypothesis does not fit into a scientific world view. At this point we have to make a decision. Should we therefore reject propositional attitude psychology, and therefore the Linguistic Relativity Hypothesis, as false and/or meaningless (eliminative materialism); or should we assume that the disciplines based on the propositional attitudes have a distinct epistemology, and therefore form an autonomous field of study? I argue for the latter.

Chapter four is an examination of the Linguistic Relativity Hypothesis in the context of the "non-scientific" approach to language and thought that has been developed in the previous chapters. The starting point of this chapter is an investigation of "radical interpretation", that is, the task of determining the thoughts of a foreign community, along with the truth-conditions of the sentences of their language. Then I review Donald Davidson's argument that this methodology entails that Whorf's Linguistic Relativity Hypothesis cannot possibly be true. However, I argue that Davidson goes too far in denying cognitive variability. There are at least two dimensions of cognitive variation that we should admit. One is due to the referential structure of language, and the other is due to the possibility of different "styles of reasoning". However, both these types of variability come about because of public socio-historical events. They are not due to the "unconscious influence" of language, as Whorf claimed. I conclude that the Linguistic Relativity Hypothesis is incoherent when viewed as a hypothesis of "hard" science, and it is false when viewed (as it should be) as a hypothesis of social science.

CHAPTER TWO - NATURAL LANGUAGE

2.1 INTRODUCTION

Whorf assumed that natural languages have a unique set of grammatical categories, and that these categories have semantic correlates. The purpose of this chapter is to examine several linguistic theories to determine whether any of them can support Whorf's assumptions.

2.2 BLOOMFIELD'S LINGUISTIC PROGRAM

Bloomfield on where to begin

Bloomfield believed that pragmatic/semantic considerations were needed to define what a natural language is, in general terms. However, at the present time linguists should forget the details of the pragmatic/semantic issues, and concentrate on structural or syntactic considerations.

Bloomfield on the description of a natural language

The structural description of a language involves three levels of description: phonology, morphology and sentence-syntax. The latter two levels are defined, in general terms, by means of semantic considerations. However, detailed semantic description is to be put on hold for future generations.

The discovery procedures

Bloomfield believed that the three levels could be characterized through the application of mechanical "discovery procedures" that can be applied with minimal semantic characterization. The discovery procedures were supposed to be linguistics' autonomous methodology. The discovery procedures were supposed to work from the bottom up.

Whorf and Bloomfield's program

Whorf wrote as if his views can be incorporated within the Bloomfieldian paradigm. He never said anything that would lead us to believe that he questioned this dominant approach. But can Whorf's views be worked in? Can a semantics be incorporated into Bloomfield's theory along the lines that he laid down?

No, his semantic theorizing is not viable. It is a mixture of referential semantics, behavioral semantics and translational semantics. The most charitable interpretation of Bloomfield's semantic theory is that it is a translational approach along the lines of Quine in Word and Object. But as Davidson and others have pointed out, translational semantics is not really semantics at all.

Chomsky's critique of Bloomfield's program

In any case, Chomsky dealt a very serious blow to the Bloomfieldian paradigm. His argument was that Bloomfield's discovery procedures would only yield a certain type of grammar, phrase structure grammar, which is actually inadequate to describe human languages. This leads to a breakdown in the discovery procedure methodology, where it is replaced with the methodology of grammatical intuitions.

Where does this leave Whorf? We found that his dependence on the Bloomfieldian paradigm was ill-placed. Can he somehow be worked into the Chomskyan paradigm? We turn to that issue in the next section.

2.3 CONCEPTUAL SEMANTICS

Chomsky's work had a profound influence on linguists. They perceived it as a license to theorize more freely. The theorizing was initially restricted to syntax, but soon incorporated semantic issues. The semantic approach that was incorporated into the Chomskyan framework was derived, in large part, from the Swiss linguist Ferdinand de Saussure. This Chomsky-Saussure synthesis is very compatible with Whorf's theoretical approach (except for the nativism characteristic of Chomsky's work). However, there are serious problems with this work. The following sections outline these points in more detail.

Saussarian Structuralism

Saussure argued that language is essentially a mapping between sounds and concepts. Consequently, language should be studied from the "inside". In particular, semantic studies should be conducted by examining internal relations within language (e.g., synonymy, hyponymy, etc.) rather than by studying the relation between language and the world.

Componential Analysis

An almost immediate consequence of Saussure's approach is that a relation of "semantic inclusion" appears to hold over linguistic expressions. In order to account for the apparent fact that some expressions "contain" the "meanings" of others, "componential analysis" evolved as a technique within the Saussarian tradition. Componential analysis postulates the existence of atomic meanings (or "semantic markers"), and views the meanings of linguistic expressions as constructed out of these atoms.

Semantics in Generative Grammar

Saussarian concepts were incorporated within the Chomskyan framework by a number of theorists, beginning in the mid-1960s.

Katz and Fodor's Theory

Jerrold Katz and Jerry Fodor produced the first such theory in 1963. However this theory fails as an adequate semantics for the same reason that Bloomfield's did. That is, it is at bottom nothing more than a translation system, and translation systems fail to meet a basic criterion of semantic theories, which is to impart knowledge adequate for the understanding of a language.

Jackendoff's Semantic Theory

The Katz-Fodor theory cannot impart knowledge adequate for the understanding of a language because it does not account for the connection of language and the world. Ray Jackendoff has recently developed a theory within the Saussure-Chomsky tradition that does address the language-world connection. Jackendoff argues that humans have internal cognitive structures that stand in a fairly isomorphic relation to linguistic structures. The elements of our cognitive structures stand in various semantic relations to the world, and linguistic structures inherit these semantic relations from the cognitive structures to which they are related. However, Jackendoff adopts a Kantian perspective, and argues that the "world" that our thoughts and language are about is not an objective, "real" world, but a "projected" world that is largely our own construction.

Jackendoff's Conceptual Semantics and Whorf's Thesis

Except for the nativism that Jackendoff inherits from Chomsky, Jackendoff's theory is highly compatible with Whorf's Linguistic Relativity Hypothesis. Stripped of the nativistic elements, Jackendoff's work can be viewed as a clearer, more detailed formulation of Whorf's theory.

Problems with Conceptual Semantics

However, there are two serious problems with Jackendoff's theory. One is that his Kantianism is poorly motivated and indefensible in any case. Secondly, Jackendoff assumes that semantic properties are essentially a matter of individual psychology. However, in the mid-1970s Hilary Putnam demonstrated that semantic properties are essentially social and indexical, and therefore a theory like Jackendoff's is inherently incapable of explaining semantic phenomena.

At this point we do not yet have a theoretical approach to language capable of explaining semantic phenomena. Consequently we do not yet have a theoretical approach capable of expressing the Linguistic Relativity Hypothesis.

2.4 TRUTH-CONDITIONAL SEMANTICS

Truth-conditional semantics was invented by Alfred Tarski as a means of precisely characterizing the relation between certain artificial languages and what they are "about". In recent years a number of theorists have argued that Tarski's methods can be applied to the study of natural language. This has come to be known as the "truth-conditional program" for the study of natural language.

Naturalism About People and Their Environments

One characteristic of the truth-conditional program is that its advocates assume a naturalistic attitude about people and their environments. That is, rather than assuming the Kantian views of Jackendoff, advocates of the truth-conditional program assume a realistic attitude similar to that held by biologists.

Linguistic Behavior as a Species of Action

Advocates of the truth-conditional program also assume that in order to view people as language users it is necessary to view them as actors. That is, linguistic episodes must be explained as "actions", which are behaviors that can be explained in terms of the subject's beliefs and desires (propositional attitudes). Grice's analysis of the conditions under which a person can "non-naturally mean" something by an utterance demonstrates the indispensability of the action framework for characterizing behavior.

Linguistic Platonism

Advocates of the truth-conditional program also argue that languages are formal systems that cannot be reduced to patterns of linguistic behavior (or, more accurately, action). Using a language is not a property of a community. Rather, language use is a two place relation between a community and a formal language.

Tarskian Truth Definitions as a Framework for Semantics

The key assumption of advocates of the truth-conditional program is that the truth-definitions of the sort that Tarski developed for artificial languages can be used as the core of an empirical theory of natural languages.

The reason why truth-definitions are appropriate comes from consideration of "interpretation" or "understanding". In the context of a general theory of action it is reasonable that we should attempt to have a theory of linguistic interpretation, knowledge of which would be sufficient for interpreting (or understanding) the linguistic actions of people who fall within the scope of the theory. Tarskian truth-definitions are exactly what is required to construct such a theory of interpretation. If the theory of understanding is understood as comprised of two components, a theory of force and a theory of meaning, a Tarskian truth-definition will meet all the demands of an adequate theory of meaning, and will do so without postulating "meanings" as entities.

Some Key Issues for the Truth-Conditional Program

It should not be assumed that all is smooth sailing for the truth-conditional program. There are a number of modifications that have to be made to Tarski's approach, and topics like intensionality and the psychological reality of grammatical rules raise thorny problems.

2.5 CONCLUSION: WHORF'S THESIS AND THE STUDY OF LANGUAGE

In summary, Whorf's Linguistic Relativity Hypothesis requires an approach to natural language that can yield a semantic theory. Bloomfield's approach cannot do this, and neither can "conceptual semantics", as practised by theorists like Ray Jackendoff. The truth-conditional program, in spite of its difficulties, is the only approach that seems likely to yield a coherent approach to semantics. Consequently, the Linguistic Relativity Hypothesis must be evaluated in the context of this theory.

CHAPTER 3 - THOUGHT

3.1 THE GEISTESWISSENSCHAFTEN

In the previous chapter it was argued that the only prospect for an adequate semantic theory (which is necessary for the evaluation of the Linguistic Relativity Hypothesis) is the truth-conditional program. This theory assumes a larger theoretical framework: the theory of action. The theory of action has been advocated as the core of the social sciences (the *Geisteswissenschaften*) by a number of theorists, but the scientific status of the theory of action is a matter of some debate. A key question is whether the social sciences are "explainable" in terms of the natural sciences. And if the answer is no, a second question is whether the action approach should be abandoned, or whether it should be pursued as a field of study that is essentially independent of the natural sciences.

Depending on how these two questions are answered, we get three different perspectives on the Linguistic Relativity

Hypothesis. (1) If social science can be integrated into natural science then the Linguistic Relativity Hypothesis must be evaluated according to standard scientific practises. (2) If integration is not possible and the social sciences are abandoned as essentially "meaningless", then the Linguistic Relativity Hypothesis must be abandoned as well. (3) If integration is not possible and the social sciences are pursued as an autonomous discipline, then the Linguistic Relativity Hypothesis must be evaluated according to methodological principles that are unique to the social sciences.

In this chapter I argue in favor of position (3). (The final chapter investigates the unique methodological principles of the social sciences and how they bear on the Linguistic Relativity Hypothesis.)

3.2 NON-PROPOSITIONAL THEORIES OF THOUGHT

However, before arguing for the autonomy of social science it should be noted that some theorists have advocated an approach to thought that is intentional, in Brentano's sense, but non-propositional. If this view is correct it spells trouble for my argument, for in chapter four when I consider methodological issues I essentially assume that thought is propositional in nature.

Advocates of non-propositional theories of thought seem to have two reasons for arguing this way. One is the adoption of a Kantian view of reality, which I reject (without too much in the way of argument). The other reason is that if thought were propositional, our cognitive processes would be monstrously complex. We would not have the time or brain-power to think the thoughts that we do, or so the critic of propositional theories of thought argues. However, this criticism makes sense only if one assumes that a theory of thought is supposed to be a description of what goes on in the brain. If one denies that, which I do in the following sections, there is no reason to reject propositional theories.

3.3 REDUCIBILITY, SUPERVENIENCE, AND PROJECTIBILITY

Now back to our first question. Is propositional attitude psychology (and thus the *Geisteswissenschaften*) "explainable" by the natural sciences.

Philosophical Behaviorism

One early attempt to answer in the positive was philosophical behaviorism, in which an attempt was made to reduce propositional attitudes to dispositions to behavior. However, this cannot possibly work because of the holistic nature of propositional attitudes. (That is, a given bit of behavioral evidence supports the attribution of a propositional attitude only on the assumption that the subject an indefinite number of other propositional attitudes.) This shows that the propositional attitude psychology does not "reduce" to any natural science in the way that, say, classical thermodynamics reduces to statistical mechanics.

The Neo-Wittgensteinian Argument

Is there any other way in which propositional attitude psychology might be "explainable" in terms of the natural sciences? Following Wittgenstein, a number of theorists argued not, claiming that explaining actions in terms of propositional attitudes was not to give a causal explanation, but merely to redescribe the action in a special vocabulary.

Reducibility and Supervenience

Donald Davidson argued that the neo-Wittgensteinian claim was wrong. As a preliminary to explaining Davidson's argument, it is important to note a distinction between "reduction" and "supervenience". One science reduces to another if each "type" (of object, process, etc.) in the reduced science is a "type" in the reducing science. Under such conditions the laws in the reduced science will be viewed as special cases of the more general laws in the reducing science.

Supervenience is a weaker relation. One science is supervenient on another if each "token" of the types in the supervening science is a token describable in the vocabulary of the more general science. However, even if supervenience holds, it does not follow that reduction will hold. Thus supervenience is indeed weaker than reducibility.

Davidson's Refutation of the Neo-Wittgensteinians

Davidson argued that propositional attitude psychology is supervenient on the sciences of the body. Furthermore, he argues, in opposition to the Wittgensteinians, that reasons can be the causes of actions. Because reasons are supervenient on physiological states, the causal laws that support the causal attribution can be couched in the vocabulary of physiology rather than propositional attitude psychology. Thus Davidson can defuse the Wittgensteinian complaint that causal explanations require causal laws, but no such laws are to be found in propositional attitude psychology.

Goodman on Projectibility

However, Davidson's argument raises other concerns. Goodman has argued that of all the imaginable predicates, we use those that are "projectible", that is, capable of supporting inductive generalizations. But according to Goodman, what makes a predicate projectible is how well it is "entrenched" in vocabulary that has supported previously successful inductions. Since the most successful inductions are those of physics, this leads to the view that the best predicates are the ones that support reduction to physics.

Projectibility and Propositional Attitudes

But this is incompatible with Davidson's view that the propositional attitudes are merely supervenient on physics. Davidson's response is that the physical predicates (i.e., the predicates of the Naturwissenschaften) form one family of mutually entrenched predicates, whereas the propositional attitude predicates (i.e., the predicates of the Geisteswissenschaften) form another family.

However, we now need an account of the principle upon which propositional attitude psychology "hangs together". The next two sections consider two such accounts. The first is "functionalism", which I interpret supporting the claim that propositional attitude psychology is related to the natural sciences, although not in the sense of reduction. However, I will reject functionalism. The other account, which I support, is Daniel Dennett's "instrumentalist" position.

3.4 FODOR'S FUNCTIONALISM

Functionalism

Functionalism is a doctrine which holds that psychological states are functional states, rather than physical states. Functionalism is weaker than ordinary reduction, but stronger than mere supervenience.

Fodor's Assumptions

Jerry Fodor is a functionalist who holds the following five assumptions.

1. Folk psychology is largely true.

2. The explanatory power of folk psychology derives from the "content" relations between propositional attitudes.
3. Scientific psychology should use the same explanatory model (actions are determined as a quasi-rational function of the agent's beliefs and desires) as that used in folk psychology.
4. Psychological states and processes are real.
5. Materialism is true.

The Mid-Seventies Theory

Fodor criticized the early version of functionalism developed in the 1960s by Putnam. In the mid-1970s Fodor developed his own version known alternatively as the Representational Theory of Mind or the Language of Thought theory. On that theory, to have a propositional attitude is to be in a particular relation to a token in an internalized language of thought, where the language of thought is a neurological code.

This theory differs from classical theoretical reduction in a number of ways, but it is similar enough in spirit that we can conclude that Fodor has "explained" propositional attitude psychology in terms of natural science, and has thus "integrated" propositional attitude psychology into the natural sciences.

The "Meaning Ain't in the Head" Fiasco

However, Fodor's theory seems subject to the same argument that was earlier employed against Jackendoff, namely, Putnam's argument that the semantic properties of the propositional attitudes are not determined by individual psychology alone.

Fodor on Contents as Functions

Fodor has attempted to defend his functionalist theory against Putnam's argument by claiming that what is in the head is not a complete determination of semantic properties, but a function (in the sense of a mathematical function) that takes a context as input and yields complete semantic properties as output.

3.5 AUTONOMOUS SOCIAL SCIENCE

Reasons for Rejecting Fodor's Functionalism

However, in spite of Fodor's defence there are two reasons for rejecting his position. First, it is possible to attribute propositional attitudes effectively even where it is obvious that there is no internal correlate (i.e., token in a language of thought). Secondly, the holistic nature of propositional attitudes causes the same problems for Fodor's theory as it does for philosophical behaviorism.

Eliminative Materialism or Autonomous Social Science?

Philosophical behaviorism fails, and now we see that Fodor's functionalism fails as well. It appears that propositional attitude psychology cannot be integrated with the natural sciences, and that cognitive psychology must be formulated without reference to "content". But what of propositional attitude psychology? Daniel Dennett has argued that it should be retained as an "autonomous discipline". It holds together (i.e., its concepts are mutually "entrenched") because of its "instrumental" character. That is, it is a normative theory based on the assumption of rationality. Given the obvious utility of propositional attitude psychology in folk explanations and epistemology, and the fact that there are no damning criticisms of Dennett's instrumentalist position, it is argued that his theory provides a justification for autonomous social science.

CHAPTER FOUR - RELATIVITY

4.1 INTRODUCTION

The strategy of this chapter will be to examine the methodology of determining the propositional attitudes of a community, and the truth-conditions of the sentences of the language used by that committee. The point of doing so will be to determine whether the methodology is consistent with the Linguistic Relativity Hypothesis.

4.2 ATTITUDES, MEANING, AND PHYSICS

The previous chapter demonstrates that attitudes and meanings cannot be reduced to physics. But can attitudes or meanings be reduced to each other?

Can Meanings be Reduced to Attitudes?

Grice has defended a program in which meanings are reduced to certain complex propositional attitudes. However, it is argued that this program is flawed for two reasons. First, it is not possible, without making linguistic attributions at the same time, to identify attitudes precisely enough to support linguistic attributions. Secondly, linguistic expressions are productive (i.e., novel utterances can be produced without limit), but it does not seem possible to develop a theory of novel attitudes that does not presuppose linguistic productivity.

Can Attitudes be Reduced to Meanings?

Although the writings of Barthes, Levi-Strauss, etc., suggest something along these lines, no serious theory of this type has been proposed.

Two Methodologies or One?

Since neither attitudes or meanings are reducible to each other, our methodology must show how both are determined on the basis of the evidence. However, can they be determined independently or must they be determined jointly? The latter course is required, because it is not possible to develop independent methodologies that can proceed without the results of the other.

Supervenience or Instrumental Alternatives?

The relation between attitudes/meanings and physics is not reduction, nor is it supervenience. Dennett's instrumental theory of intentionality supports an even weaker view: that the intentional "sciences" is an instrumental alternative to the natural sciences.

4.3 RADICAL INTERPRETATION

"Radical interpretation" is a term that has been coined to refer to the methodological problem of determining a person's attitudes and meanings from physical evidence.

Davidson on Radical Interpretation

Davidson has developed a method a radical interpretation based on the assumption that the social scientist can directly observe cases of a subject holding a sentence true (under certain environmental conditions).

Lewis on Radical Interpretation

David Lewis has pointed out that the central issue in discussions of radical interpretation is not that of finding a practical method, but of precisely determining the constraints on the attributions of attitudes and meanings. Consideration of Lewis' work shows that "rationality" and "causality" are not complementary constraints. Rather, "causality" should be rejected as a constraint.

4.4 DAVIDSON ON CONCEPTUAL RELATIVITY

Davidson has argued that the methodology of radical interpretation entails that conceptual relativity is impossible. He argues first of all that conceptual incommensurability is best understood as failure of translation. He then argues that neither total or partial failure of translation is possible.

Total Failure of Translation

Davidson claims that we can have two incommensurable conceptual schemes only if both are largely true, but they are incommensurable. However, he goes on to claim that the concept of truth implies translatability. I object to his argument, but something very similar can be defended.

Partial Failure of Translation

The principle of charity entails that translational failures must be very small, for significant failure undermines the coherence of what has been successfully translated.

Davidson concludes that conceptual relativity is a false doctrine. He assumes that his arguments apply to Whorf's Linguistic Relativity Hypothesis.

4.5 RELATIVITY OF REFERENCE

Bruce Aune thinks that Davidson has overstated the case against conceptual relativity because he has an incorrect account of translation.

Aune on Translation

Davidson believes that translation can be nothing more than the correlation of sentences. By employing an argument of Goodman's, Aune defends the view that translation also involves the correlation of terms.

Aune on Conceptual Schemes

Aune also defends a view of conceptual schemes that is different than Davidson's. He argues that conceptual schemes are defined by systems of predicates; that these are consciously introduced into communities by advocates; and that there can be several competing conceptual schemes within a single speech community.

Comments on Aune's Argument

There are two issues surrounding Aune's argument. First, he has not developed a satisfactory theory of what is involved in introducing a new conceptual scheme into a community. This problem will be ignored in what follows. Secondly, he assumes that it is possible to objectively determine what predicates a speaker possesses. However, Davidson and others have questioned this second assumption.

Davidson on Reference

Davidson argues that the empirical base of radical interpretation is a set of judgements about which sentences the subject holds-true. Predicates, names, etc., do not have any empirical reality, rather, they are merely the theoretician's invention as he attempts to construct a finite theory to account for the evidence available to him.

The Causal Theory of Reference

It might be thought that the "causal theory of reference" establishes that the semantic properties of sub-sentential expressions can be empirically determined.

The Inscrutability of Reference

However, Hartry Field has shown that reference is inscrutable even if one does adopt a causal theory.

The Reality of Reference - A Defence

In spite of Field's argument, I maintain that reference is just as "real" as the semantics of sentences. To argue, as Davidson does, that sentences have empirical reality whereas the clauses of a truth-definition governing reference do not, is to fall back into a pre-Quinian philosophy of science that assumes that theories can be verified a sentence at a time.

Consequently, a theory along the lines of Aune's is possible, and thus it is possible to recognize a greater degree of cognitive variation than Davidson assumes.

4.6 RELATIVITY OF REASONING

Another possible dimension of variation that Davidson overlooks is variation in the way people reason; that is, in how they determine new beliefs and desires on the basis of previous ones and new evidence.

A fruitful way of approaching this question is by considering alternate theories for explaining differences in the way that people reason.

The Performance Error Theory

One theory is that all people are essentially rational. Therefore when two people differ in their reasoning, at least one is making performance errors (i.e., memory lapses, slips of the tongue, etc.). Note that such a theory would not produce an interesting version of the Linguistic Relativity Hypothesis.

The Theory of Systematic Irrationality

A second theory is that at least one person is systematically irrational. Note that this option is not open to you if you believe that rationality is constitutive of having propositional attitudes. Note that such a theory does support a version of the Linguistic Relativity Hypothesis, but a rather uninteresting one, for language is seen as contributing to the degree of irrationality, and thus is viewed as entirely negative in nature.

The Theory of Relativity of Rationality

A third theory is that rationality itself is relative to a culture. This theory allows for a very powerful version of the Linguistic Relativity Hypothesis.

The Socio-Historical Emergence Theory of Rationality

A fourth and final theory is that rationality is partially culturally determined, and partly universal. This theory also allows for a fairly powerful version of the Linguistic Relativity Hypothesis.

The task at hand is to determine which, if any, of these four theories is correct.

Are Standards of Rationality A Priori?

The first two theories entail that standards of rationality are given a priori. However, I agree with the vast majority of twentieth century writers, who reject the idea of a priori knowledge or standards. The concept of the a priori was invented in order to account for aspects of our knowledge that go beyond the empirical evidence. However, the pragmatists' theory of the organism's contribution to knowledge is far preferable to the theory of a priori knowledge. Consequently, our choice is between the third and the fourth theory described above.

Putnam on Why Reason Can't be Naturalized

Putnam has effectively argued that reason cannot be completely naturalized, for that would leave us in an untenable position that he calls "cultural imperialism". This means that the third theory cannot be true, and thus we a left with the socio-historical emergence theory of rationality.

Rationality and Language

The socio-historical emergence theory of rationality has been implicitly adopted by a large number of theorists. Using this theory it is possible to describe significant conceptual differences between (and within) communities; but these differences always have a public history. This view of conceptual variation is incompatible with Whorf's view of conceptual variation that is a-historical and unconsciously caused.

4.7 CONCLUSION

The proper theoretical context for Whorf's Linguistic Relativity Hypothesis is a theory of action which incorporates a Tarskian truth theory in order to make sense of linguistic actions. However, in considering the methodology associated with such a theory, we have seen that the amount of conceptual variation between communities that can be postulated in such a theory is limited, and that the causes of the variation are always non-linguistic.